(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 June 2002 (13.06.2002)

PCT

(10) International Publication Number
WO 02/46467 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/IB01/02811

(22) International Filing Date: 7 December 2001 (07.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/254,090	8 December 2000 (08.12.2000)	US
10/007,926	7 December 2001 (07.12.2001)	US

(71) Applicant *(for all designated States except US)*: **IPSOGEN** [FR/FR]; 232 Boulevard Sainte-Marguerite, F-13273 Marseille Cedex 09 (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BERTUCCI, François** [FR/FR]; Le Sully, Parc de la Cadenelle, 122, rue du Commandant Rolland, F-13008 Marseille (FR). **HOULGATTE, Rémi** [FR/FR]; 31, Boulevard Lombard, F-13015 Marseille (FR). **BIRNBAUM, Daniel** [FR/FR]; 9, rue Baldacchini, La Croix-du-Sud, F-13009 Marseille (FR). **NGUYEN, Catherine** [FR/FR]; 8 Boulevard de la Kabylie, F-13016 Marseille (FR). **VIENS, Patrice** [FR/FR]; La Palmeraie Borely, Entrée A, 70, avenue d'Haïfa, F-13008 Marseille (FR). **FERT, Vincent** [FR/FR]; 971, chemin du Tardinaou, F-13190 Allauch (FR).

(74) Agents: **BREESE, Pierre** et al.; Breese-Majerowicz, 3, avenue de l'Opéra, F-75001 Paris (FR).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/46467 A2

(54) Title: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS USING ARRAYS OF CANDIDATE GENES

(57) Abstract: The invention relates to a polynucleotide library useful in the molecular characterization of a carcinoma, the library including a pool of polynucleotide sequences of subsequences thereof wherein the sequences of subsequences are overpressed in tumor cells, further wherein the sequences of subsequences correspond substantially to any of the polynucleotide sequences set forth in any of SEQ ID NOS: 1-468 or the complement thereof. The invention relates also to polynucleotides arrays useful to differentiate tumor cells from normal cells comprising combinations of selected immobilized polynucleotide sequences sets.

GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS USING ARRAYS OF CANDIDATE GENES

This invention relates to polynucleotide analysis and, in particular, to polynucleotide expression profiling of carcinomas using arrays of candidate polynucleotides.

Pathologists and clinicians in charge of the management of breast cancer patients are facing two major problems, namely the extensive heterogeneity of the disease and the lack of factors - among conventional histological and clinical features - predicting with reliability the evolution of the disease and its sensitivity to cancer therapies. Breast tumors of the same apparent prognostic type vary widely in their responsiveness to therapy and consequent survival of the patient. New prognostic and predictive factors are needed to allow an individualization of therapy for each patient.

Great hope is currently being placed on molecular studies, which address the problem in a global fashion. Methods such as cytogenetics, comparative genomic hybridization, and whole-genome allelotyping have addressed the issue at the genome level. Currently, the modifications that take place in human tumors at the level of transcription can also be studied in a large, unprecedented scale, using new methods such as cDNA arrays that allow quantitative measurement of the mRNA expression levels of many genes simultaneously. Thus, it would be advantageous to provide a means to assess the capacity of cDNA array testing in clinical practice to better classify an heterogeneous cancer into tumor subtypes with more homogeneous clinical outcomes, and to identify new potential prognostic factors and therapeutics targets.

The invention relates to a polynucleotide library useful in the molecular characterization of a carcinoma, the library including a pool of polynucleotide sequences or subsequences thereof wherein the sequences or subsequences are either underexpressed or overpressed in tumor cells, further wherein the sequences or subsequences correspond substantially to any of the polynucleotide sequences set forth in any of SEQ ID NOS: 1 - 468 or the complement thereof.

Fig. 1 shows an example of differential gene expression between normal breast tissue (NB) and breast tumor samples.

Fig. 2 is a representation of expression levels of 176 genes in normal breast tissue (NB) and 34 samples of breast carcinoma.

Fig. 3 is prognostic classification of breast cancer by gene expression profiling.

Fig. 4 shows the correlation of GATA3 expression with ER phenotype.

In the context of this disclosure, a number of terms shall be utilized.

The term "polynucleotide" refers to a polymer of RNA or DNA that is single-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

The term "subsequence" refers to a sequence of nucleic acids that comprises a part of a longer sequence of nucleic acids.

The term "immobilized on a support" means bound directly or indirectly thereto including attachment by

covalent binding, hydrogen bonding, ionic interaction, hydrophobic interaction or otherwise.

Breast cancer is characterized by an important histoclinical heterogeneity that currently hampers the selection of the most appropriate treatment for each case. This problem could be solved by the identification of new parameters that better predict the natural history of the disease and its sensitivity to treatment. An important object of the present invention relates to a large-scale molecular characterization of breast cancer that could help in prediction, prognosis and cancer treatment.

An important aspect of the invention relates to the use of cDNA arrays, which allows to quantitative study mRNA expression levels of 188 candidate genes in 34 consecutive primary breast carcinomas along three directions: comparison of tumor samples, correlations of molecular data with conventional histoclinical prognostic features and gene correlations. The experimentation evidenced extensive heterogeneity of breast tumors at the transcriptional level. Hierarchical clustering algorithm identified two molecularly distinct subgroups of tumors characterized by a different clinical outcome after chemotherapy. This outcome could not have been predicted by the commonly used histoclinical parameters. No correlation was found with the age of patients, tumor size, histological type and grade. However, expression of genes was differential in tumors with lymph node metastasis and according to the estrogen receptor status; ERBB2 expression was strongly correlated with the lymph node status ($p \leq 0.0001$) and that of GATA3 with the presence of estrogen receptors ($p \leq 0.001$). Thus, experimental results identified new ways to group tumors according to outcome and new potential targets of carcinogenesis. They show that the systematic use of cDNA

array testing holds great promise to improve the classification of breast cancer in terms of prognosis and chemosensitivity and to provide new potential therapeutic targets.

DNA arrays consist of large numbers of DNA molecules spotted in a systematic order on a solid support or substrate such as a nylon membrane, glass slide, glass beads or a silicon chip. Depending on the size of each DNA spot on the array, DNA arrays can be categorized as microarrays (each DNA spot has a diameter less than 250 microns) and macroarrays (spot diameter is grater than 300 microns). When the solid substrate used is small in size, arrays are also referred to as DNA chips. Depending on the spotting technique used, the number of spots on a glass microarray can range from hundreds to thousands.

DNA microarrays have serve a variety of purposes, including, gene expression profiling, de novo gene sequencing, gene mutation analysis, gene mapping and genotyping. cDNA microarrays are printed with distinct cDNA clones isolated from cDNA libraries. Therefore, each spot represents an expressed gene, since it is derived from a distinct mRNA.

Typically, a method of monitoring gene expression involves providing (1) providing a pool of sample polynucleotides comprising RNA transcript(s) of one or more target gene(s) or nucleic acids derived from the RNA transcript(s); (2) reacting, such as hybridizing the sample polynucleotide to an array of probes (for example, polynucleotides obtained from a polynucleotide library) (including control probes) and (3) detecting the reacted/hybridized polynucleotides. Detection can also involve calculating/quantifying a relative expression (transcription) level.

The present invention concerns a polynucleotide library useful in the molecular characterization of a carcinoma, said library comprising a pool of polynucleotide sequences or subsequences thereof wherein said sequences or subsequences are either underexpressed or overpressed in tumor cells, further wherein said sequences or subsequences correspond substantially to any of the polynucleotide sequences set forth in any of SEQ ID Nos: 1 - 468 in annex or the complement thereof.

Obviously, sequences having a great degree of homology with the above sequences could also been used to realize the molecular caracterization of the invention, namely when those sequences present one or a few punctual mutations when compared with anyone of sequences SEQ ID Nos: 1 - 468.

The invention concerns a polynucleotide library useful in the molecular characterization of a carcinoma, said library comprising a pool of polynucleotide sequences or subsequences thereof wherein said sequences or subsequences are overpressed in tumor cells, further wherein said sequences or subsequences correspond substantially to any of the polynucleotide sequences set forth in any of SEQ ID NOS: 1 - 249 (Here, these SEQ ID N° refer to old SEQ ID N° 1-249 in priority document, the correlation table 10 allows to identify these sequences in the sequence listing of the present application in annex) or the complement thereof

Preferably the pool of polynucleotide sequences or subsequences correspond substantially to the polynucleotide sequences set forth in any of SEQ ID NOS: 1 - 247 (Here, these SEQ ID N° refer to old SEQ ID N° 1-247 in priority document, the correlation table 10 allows to identify these sequences in the sequence listing of the present application

in annex); further wherein said sequences are useful in differentiating a normal cell from a cancer cell.

The invention relates also to a polynucleotide library wherein the pool of polynucleotide sequences or subsequences correspond substantially to the polynucleotide sequences set forth in any of SEQ ID NOS: 1 - 242 (Here, these SEQ ID N° refer to old SEQ ID N° 1-242 in priority document, the correlation table 10 allows to identify these sequences in the sequence listing of the present application in annex); wherein said sequences are useful in detecting a hormone sensitive tumor cell, or wherein said sequences are useful in differentiating a tumor with lymph nodes from a tumor without lymph nodes.

The invention relates also to a polynucleotide library wherein the pool of polynucleotide sequences or subsequences correspond substantially to the polynucleotide sequences set forth in any of SEQ ID NOS: 1 - 224; (Here, these SEQ ID N° refer to old SEQ ID N° 1-224 in priority document, the correlation table 10 allows to identify these sequences in the sequence listing of the present application in annex) wherein said sequences are useful in differentiating tetracycline-sensitive tumors from tetracycline-insensitive tumors.

The invention relates also to any polynucleotide library as previously described wherein said polynucleotides are immobilized on a solid support in order to form a polynucleotide array.

Preferably the support is selected from the group consisting of a nylon membrane, glass slide, glass beads, or a silicon chip.

The invention concerns also a method for detecting differentially expressed polynucleotide sequences which are correlated with a cancer, said method comprising:

a) obtaining a polynucleotide sample from a patient; and

b) reacting the sample polynucleotide obtained in step (a) with a probe immobilized on a solid support wherein said probe comprises any of the polynucleotide sequences of the libraries previously described or an expression product encoded by any of the polynucleotide sequences of said libraries and

c) detecting the reaction product of step (b).

The invention relates also to a such method for detecting differentially expressed polynucleotide sequences of the invention wherein the amount of reaction product of step (c) is compared to a control sample.

Preferably the polynucleotide sample isolated for, the sample is RNA or mRNA.

Preferably the polynucleotide sample is cDNA obtained by reverse transcription of the mRNA.

In a prefered embodiment the method for detecting differentially expressed polynucleotide sequences, the step (b) comprises a hybridization of the sample RNA with the labeled probe.

The method for detecting differentially expressed polynucleotide sequences is used for detecting, diagnosing, staging, monitoring, prognosticating, preventing or treating conditions associated with cancer, and namelly breast cancer.

The method for detecting differentially expressed polynucleotide sequences is particular useful wherein the product encoded by any of the polynucleotide sequences or

subsequences is involved in a receptor-ligand reaction on which detection is based.

The invention relates also to a method for screening an anti-tumor agent comprising the method for detecting differentially expressed polynucleotide sequences previously described wherein the sample has been treated with the anti-tumor agent to be screened.

Le label used to label polynucleotide samples is selected from the group consisting of radioactive, colorimetric, enzymatic, molecular amplification, bioluminescent or fluorescent label.

Yhe invention also relates to a library of polynucleotides comprising a population of polynucleotide sequences overexpressed or underexpresses in cells derived from a tumor selected from SEQ ID NO :1 to SEQ ID NO :249 and their respective complements. (Here, these SEQ ID N° refer to old SEQ ID N° 1-249 in priority document, the correlation table 10 allows to identify these sequences in the sequence listing of the present application in annex).

In a particular embodiment the invention relates to polynucleotide sequences: SEQ ID No : 1 ; SEQ ID No : 5 ; SEQ ID No : 8 ; SEQ ID No : 9 ; SEQ ID No : 28 ; SEQ ID No : 29 ; SEQ ID No : 30 ; SEQ ID No : 31 ; SEQ ID No : 32 ; SEQ ID No : 45 ; SEQ ID No : 46 ; SEQ ID No : 52 ; SEQ ID No : 54 ; SEQ ID No : 63 ; SEQ ID No : 64 ; SEQ ID No : 81 ; SEQ ID No : 82 ; SEQ ID No : 87 ; SEQ ID No : 88 ; SEQ ID No : 101 ; SEQ ID No : 102 ; SEQ ID No : 103 ; SEQ ID No : 104 ; SEQ ID No : 105 ; SEQ ID No : 107 ; SEQ ID No : 113 ; SEQ ID No : 114 ; SEQ ID No : 115 ; SEQ ID No : 116 ; SEQ ID No : 127 ; SEQ ID No : 128 ; SEQ ID No : 131 ; SEQ ID No : 139 ; SEQ ID No : 140 ; SEQ ID No : 142 ; SEQ ID No : 150 ; SEQ ID No : 151 ; SEQ ID No : 154 ; SEQ ID No : 156 ; SEQ ID

No : 160 ; SEQ ID No : 161 ; SEQ ID No : 162 ; SEQ ID No : 177 ; SEQ ID No : 178 ; SEQ ID No : 194 ; SEQ ID No : 195 ; SEQ ID No : 227 ; SEQ ID No : 228 ; SEQ ID No : 229 ; SEQ ID No : 231 ; SEQ ID No : 233 ; SEQ ID No : 243 ; SEQ ID No : 244 ; SEQ ID No : 245 ; SEQ ID No : 246 ; SEQ ID No : 247, (Here, these SEQ ID N° refer to old SEQ ID N° presented on table 5 in priority document, the correlation table 10 allows to identify these sequences in the sequence listing of the present application in annex), which distinguish a healthy person from a person with cancer.

Preferably the invention relates to polynucleotide sequences: SEQ ID No : 1 ; SEQ ID No : 5 ; SEQ ID No : 102 ; SEQ ID No : 103 ; SEQ ID No : 107 ; SEQ ID No : 229 ; SEQ ID No : 45 ; SEQ ID No : 46; SEQ ID No : 243 ; SEQ ID No : 244; SEQ ID No : 245 ; SEQ ID No : 246 ; SEQ ID No : 247 (Here, these SEQ ID N° refer to old SEQ ID N° presented on table 6 in priority document, the correlation table 10 allows to identify these sequences in the sequence listing of the present application in annex), which distinguish a healthy person from a person with cancer.

In another particular embodiment the invention relates to polynucleotide sequences: SEQ ID No : 2 ; SEQ ID No : 3; SEQ ID No : 4 ; SEQ ID No : 5 ; SEQ ID No : 6 ; SEQ ID No : 7 ; SEQ ID No : 8 ; SEQ ID No : 9 ; SEQ ID No : 10 ; SEQ ID No : 11 ; SEQ ID No : 12 ; SEQ ID No : 13 ; SEQ ID No : 14 ; SEQ ID No : 15 ; SEQ ID No : 16 ; SEQ ID No : 17 ; SEQ ID No : 18 ; SEQ ID No : 19 ; SEQ ID No : 20 ; SEQ ID No : 21 ; SEQ ID No : 22 ; SEQ ID No : 23 ; ; SEQ ID No : 24 ; SEQ ID No : 25 ; SEQ ID No : 26 ; SEQ ID No : 27 ; SEQ ID No : 221 ; SEQ ID No : 222 ; SEQ ID No : 223 ; SEQ ID No : 241 ; SEQ ID No : 242 (Here, these SEQ ID N° refer to old SEQ ID N° presented on table 7 in priority document, the correlation table 10 allows to identify these sequences in the sequence listing of the present application in annex) which detect hormone sensitive tumors.

Preferably the invention relates to polynucleotide sequences SEQ ID No : 1; SEQ ID No : 2 SEQ ID No : 3; SEQ ID No : 4; SEQ ID No : 5; SEQ ID No : 221; SEQ ID No : 222 ; SEQ ID No : 15; SEQ ID No : 16; SEQ ID No : 17; SEQ ID No : 18 ; SEQ ID No : 19; SEQ ID No : 20 ; SEQ ID No : 21; SEQ ID No : 22 ; SEQ ID No : 241; SEQ ID No : 242 (Here, these SEQ ID N° refer to old SEQ ID N° presented on table 8 in priority document, the correlation table 10 allows to identify these sequences in the sequence listing of the present application in annex), which detect hormone sensitive tumors.

In another particular embodiment the invention relates to polynucleotide sequences: SEQ ID No : 1 ; SEQ ID No : 3 ; SEQ ID No : 4 ; SEQ ID No : 19 ; SEQ ID No : 20 ; SEQ ID No : 2 1; SEQ ID No : 22 ; SEQ ID No : 23 ; SEQ ID No : 26 ; SEQ ID No : 27 ; SEQ ID No : 28 ; SEQ ID No : 29 ; SEQ ID No : 30 ; SEQ ID No : 31 ; SEQ ID No : 32 ; SEQ ID No : 33 ; SEQ ID No : 34 ; SEQ ID No : 35 ; SEQ ID No : 36; SEQ ID No : 37; SEQ ID No : 38; SEQ ID No : 39; SEQ ID No : 40 ; SEQ ID No : 41 ; SEQ ID No : 42 ; SEQ ID No : 43 ; SEQ ID No : 44 ; SEQ ID No : 221 ; SEQ ID No : 222 ; SEQ ID No : 233 ; SEQ ID No : 241 ; SEQ ID No : 242 (Here, these SEQ ID N° refer to old SEQ ID N° presented on table 8 in priority document, the correlation table 10 allows to identify these sequences in the sequence listing of the present application in annex), which distinguish tumors with lymphe node from tumors with no lymphe node.

Preferably the invention relates to polynucleotide sequences : SEQ ID No : 1 ; SEQ ID No : 21 ; SEQ ID No : 22 ; SEQ ID No : 28; ; SEQ ID No : 29 ; SEQ ID No : 29 ; SEQ ID No : 31 ; SEQ ID No : 32 ; SEQ ID No : 19 ; SEQ ID No : 20 ; SEQ ID No : 26 ; SEQ ID No : 27 ; SEQ ID No : 37 ; SEQ ID No : 38 ; SEQ ID No : 39 ; SEQ ID No : 241 ; SEQ ID No : 241, (Here, these SEQ ID N° refer to old SEQ ID N° presented on table 8 in priority document, the correlation table 10 allows to

identify these sequences in the sequence listing of the present application in annex), which distinguish tumors with lymphe node from tumors with no lymphe node.

In another particular embodiment the invention relates to polynucleotide sequences: SEQ ID No : 1 ; SEQ ID No : 2 ; SEQ ID No : 6 ; SEQ ID No : 7 ; SEQ ID No : 8 ; SEQ ID No : 9 ; SEQ ID No : 10 ; SEQ ID No : 11 ; SEQ ID No : 13 ; SEQ ID No : 14 ; SEQ ID No : 19 ; SEQ ID No : 20 ; SEQ ID No : 21 ; SEQ ID No : 22 ; SEQ ID No : 23 ; SEQ ID No : 35 ; SEQ ID No : 36 ; ; SEQ ID No : 37 ; SEQ ID No : 56 ; SEQ ID No : 57 ; SEQ ID No : 74 ; SEQ ID No : 75 ; SEQ ID No : 102 ; SEQ ID No : 104 ; SEQ ID No : 107 ; SEQ ID No : 108 ; SEQ ID No : 109 ; SEQ ID No : 118 ; SEQ ID No : 119 ; ; SEQ ID No : 136 ; SEQ ID No : 213 ; SEQ ID No : 214 ; SEQ ID No : 215 ; SEQ ID No : 223 ; SEQ ID No : 224 (Here, these SEQ ID N° refer to old SEQ ID N° presented on table 11 in priority document, the correlation table 10 allows to identify these sequences in the sequence listing of the present application in annex) which distinguish tumors sensitive to antracycline from tumors unsensitive to antracycline.

The invention relates also to a method of detecting differentially expressed genes correlated with a cancer comprising detecting at least one library of polynucleotide sequences as above defined or of products encoded by said library in a sample obtained from a patient.

A particular embodiment of the invention relates to a polynucleotide library of corresponding substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets 1 to set 212 as defined in table 4

The invention relates obviously to polynucleotide libraries comprising at least one polynucleotide selected among those included in at least 50%, preferably 75% and more preferably 100% of said predefined sets, allowing to obtain a discriminating gene pattern, namely to distinguish between normal patients and patients suffering from tumor pathology, between patients having an hormone sensitive tumor and patients having an hormone resistant tumor, between patients having a tumor with lymph nodes from patients having a tumor without lymph nodes, between patients having an antracycline-sensitive tumor from patients having an antracycline-insensitive tumor and between patients having good prognosis primary breast tumors and patients having poor prognosis primary breast tumors.

Polynucleotide sequences library useful for the realization of the invention can comprise also any sequence comprised between 3'end and 5'end of each polynucléotide sequence set as defined in table 4, allowing the complete detection of the implicated genes.

The invention relates also to a polynucleotide library useful to differentiate a normal cell from a cancer cell wherein the pool of polynucleotide sequences or subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets indicated on table 5, useful in differentiating a normal cell from a cancer cell.

Preferably the polynucleotide library useful to differentiate a normal cell from a cancer cell correspond

substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets indicated on table 5A, and of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets indicated in table 5B.

The detection of an overexpression of genes identified with sets of polynucleotides sequences defined on table 5A, together with detection of an underexpression of genes identified with sets of polynucleotides sequences defined in table 5B allows to distinguish between normal patients, and patients suffering from tumor pathology.

The invention relates also to a polynucleotide library useful to detect a hormone sensitive tumor cell wherein the pool of polynucleotide sequences or subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 6

Preferably the polynucleotide library useful to detect a hormone sensitive tumor cell correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 6A together with at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 6B.

The detection of an overexpression of genes identified with sets of polynucleotides sequences defined on table 6A, together with detection of an underexpression of

genes identified with sets of polynucleotides sequences defined in table 6B allows to distinguish between patients having an hormone sensitive tumor and patients having an hormone resistant tumor.

The invention concerns also a polynucleotide library useful to differentiate a tumor with lymph nodes from a tumor without lymph nodes wherein the pool of polynucleotide sequences or subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 7.

Preferably, the polynucleotide library useful to differentiate a tumor with lymph nodes from a tumor without lymph nodes correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 7A together with at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 7B.

The detection of an overexpression of genes identified with sets of polynucleotides sequences defined on table 7A, together with detection of an underexpression of genes identified with sets of polynucleotides sequences defined in table 7B allows to distinguish between patients having a tumor with lymph nodes from patients having a tumor without lymph nodes.

The invention concerns also a polynucleotide library useful to differentiate antracycline-sensitive tumors

from antracycline-insensitive tumors wherein the pool of polynucleotide sequences or subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 8.

Preferably, the polynucleotide library useful to differentiate antracycline-sensitive tumors from antracycline-insensitive tumors correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 8A together with at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 8B.

The detection of an overexpression of genes identified with sets of polynucleotides sequences defined on table 8A, together with detection of an underexpression of genes identified with sets of polynucleotides sequences defined in table 8B allows to distinguish between patients having an antracycline-sensitive tumor from patients having an antracycline-insensitive tumor.

The invention concerns also a polynucleotide library useful to classify good and poor prognosis primary breast tumors wherein the pool of polynucleotide sequences or subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 9.

Preferably, the polynucleotide library useful to classify good and poor prognosis primary breast tumors correspond substantially to any combination of at least one

polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 9A together with at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 9B.

The detection of an overexpression of genes identified with sets of polynucleotides sequences defined on table 9A, together with detection of an underexpression of genes identified with sets of polynucleotides sequences defined in table 9B allows to classify patients having good and poor prognosis primary breast tumors.

In a preferred embodiment, the tumor cell presenting underexpressed or overpressed sequences from the polynucleotide library of the invention are breast tumor cells.

In a particular embodiment the polynucleotides of the polynucleotide library of the present invention are immobilized on a solid support in order to form a polynucleotide array, and said solid support is selected from the group consisting of a nylon membrane, nitrocellulose membrane, glass slide, glass beads, membranes on glass support or a silicon chip.

Another object of the present invention concerns a polynucleotide array useful for prognosis or diagnostic of tumor comprising at least one immobilized polynucleotide library set as previously defined.

Then the invention concerns a polynucleotide array useful to differentiate a normal cell from a cancer cell comprising any combination of at least one polynucleotide sequence selected among those included in each

one of predefined polynucleotide sequences sets indicated on table 5, useful in differentiating a normal cell from a cancer cell.

Preferably the polynucleotide array useful to differentiate a normal cell from a cancer cell bears any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets indicated on table 5A, and of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets indicated in table 5B.

The invention relates also to a polynucleotide array useful to detect a hormone sensitive tumor cell comprising any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 6

Preferably the polynucleotide array useful to detect a hormone sensitive tumor cell bears any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 6A together with at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 6B.

The invention concerns also a polynucleotide array useful to differentiate a tumor with lymph nodes from a tumor without lymph nodes comprising any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 7.

Preferably, the polynucleotide array useful to differentiate a tumor with lymph nodes from a tumor without lymph nodes bears any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 7A together with at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 7B.

The invention concerns also a polynucleotide array useful to differentiate antracycline-sensitive tumors from antracycline-insensitive tumors comprising any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 8.

Preferably, the polynucleotide array useful to differentiate antracycline-sensitive tumors from antracycline-insensitive tumors bears any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 8A together with at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 8B.

The invention concerns also a polynucleotide array useful to classify good and poor prognosis primary breast tumors comprising any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 9.

Preferably, the polynucleotide array useful to classify good and poor prognosis primary breast tumors bears any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 9A together with at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 9B.

The present invention concerns also a method for detecting differentially expressed polynucleotide sequences that are correlated with a cancer, said method comprising:

a) obtaining a polynucleotide sample from a patient; and

b) reacting the sample polynucleotide obtained in step (a) with a probe immobilized on a solid support wherein said probe comprises any of the polynucleotide sequences of the libraries previously defined or an expression product encoded by any of the polynucleotide sequences of the libraries previously defined

c) detecting the reaction product of step (b).

Preferably, the polynucléotide sample obtained at step (a) is labeled before its reaction at step (b) with the probe immobilized on a solid support.

The label of the polynucléotide sample is selected from the group consisting of radioactive, colorimetric, enzymatic, molecular amplification, bioluminescent or fluorescent.

In a particular embodiment the reaction product of step (c) is quantified by further comparison of said reaction product to a control sample.

In a first embodiment, the polynucleotide sample isolated from the patient and obtained at step (a) is either RNA or mRNA.

In another embodiment the polynucleotide sample isolated from the patient is cDNA is obtained by reverse transcription of the mRNA.

Preferably the reaction step (b) of the method for detecting differentially expressed polynucleotide sequences comprises a hybridization of the sample RNA issued from patient with the probe.

Preferably the sample RNA is labeled before hybridization with the probe and the label is selected from the group consisting of radioactive, colorimetric, enzymatic, molecular amplification, bioluminescent or fluorescent.

This method for detecting differentially expressed polynucleotide sequences is particularly useful for detecting, diagnosing, staging, monitoring, prognosticating, preventing or treating conditions associated with cancer, and particularly breast cancer.

The method for detecting differentially expressed polynucleotide sequences is also particularly useful when the product encoded by any of the polynucleotide sequences or subsequences set is involved in a receptor-ligand reaction on which detection is based.

The present invention is also related with a method for screening an anti-tumor agent comprising the method the above-depicted method for detecting differentially expressed polynucleotide sequences wherein the sample has been treated with the anti-tumor agent to be screened.

In a particular embodiment the method for screening an anti-tumor agent comprises detecting polynucleotide sequences reacting with at least one library of polynucleotides or polynucleotide sequences set as previously defined or of products encoded by said library in a sample obtained from a patient.

The invention is illustrated by examples detailed below related to particular experimental results obtained with selected libraries of polypeptides useful to identify and distinguish tumor samples from normal ones.

Tumor samples and RNA extraction

To avoid any bias of selection as to the type and size of the tumors, the RNAs to be tested were prepared from unselected samples. Samples of primary invasive breast carcinomas were collected from 34 patients undergoing surgery at the Institute Paoli-Calmette. After surgical resection, the tumors were macrodissected: a section was taken for the pathologist's diagnosis and an adjacent piece was quickly frozen in liquid nitrogen for molecular analyses. The median age of patients at the time of diagnosis was 55 years (range 39, 83) and most of them were post-menopausal. Tumors were classified according to the WHO histological typing of breast tumors in: 29 ductal carcinomas, 2 lobular carcinomas, 1 mixed ductal and lobular carcinoma, and 2 medullar carcinomas. They had various sizes, inferior or equal to 20 mm (n = 13), between 20 and 50 mm (n = 18) or superior to 50 mm (n = 3), axillary's lymph node status (negative: 19 tumors, positive: 15 tumors), SBR grading (I: 3 tumors, II: 20 tumors, III: 10 tumors, not evaluable: 1 tumor), and estrogen receptor status (ER) evaluated by

immunohistochemical assay (23 ER-positive, 11 ER-negative). ER positivity cutoff value was 10%. Adjuvant treatment with radiotherapy and when necessary multi-agent anthracyclin-based chemotherapy (n = 16) was given to patients according to local practice.

Total RNA was extracted from tumor samples by standard methods (43). Total RNA from normal breast tissue was obtained from Clontech (Palo Alto, CA): RNA was isolated from 8 tissue specimens from Caucasian females, age range 23 - 47. RNA integrity was controlled by denaturing formaldehyde agarose gel electrophoresis and Northern blots using a 28S-specific oligonucleotide.

cDNA arrays preparation

Gene expression was analyzed by hybridization of arrays with radioactive probes. The arrays contained PCR products of 5 control clones, and 180 IMAGE human cDNA clones selected with practical criteria (3' sequence of mRNA, same cloning vector, host bacteria and insert size). This represented 176 genes (4 genes were represented by 2 different clones): 121 with proven or putative implication in cancer and 55 implicated in immune reactions (the list is available on the web site: http:/tagc.univ-mrs.fr/pub/Cancer/). Their identity was verified by 5' tag-sequencing of plasmid DNA and comparison with sequences in the EST (dbEST) and nucleotide (GenBank) databases at the NCBI. Identity was confirmed for all but 14 clones without significant gene similarity, which were referenced by their GenBank accession number. The control clones were: Arabidopsis thaliana cytochrome c554 gene (used for hybridization signal normalization), 3 poly(A) sequences of different sizes and the vector pT7T3D (negative controls).

PCR amplification, purification and robotical spotting of PCR products onto Hybond-N+ membranes (Amersham) were done according to described protocols (4). All PCR products were spotted in duplicate. For normalization purpose, the c554 gene was spotted 96-fold scattered over the whole membrane.

cDNA array hybridizations

Hybridizations were done successively with a vector oligonucleotide (to precisely determine the amount of target DNA accessible to hybridization in each spot), then after stripping of vector probe, with complex probes made from the RNAs (4). Each complex probe was hybridized to a distinct filter. Probes were prepared from total RNA with an excess of oligo(dT25) to saturate the poly(A) tails of the messengers, and to insure that the reverse transcribed product did not contain long poly(T) sequences. A precise amount of c554 mRNA was added to the total RNA before labeling to allow normalization of the data.

Five ng of total RNA (~100ng of mRNA) from tissue samples were used for each labeling. Probe preparation and hybridization of the membranes were done according to known procedures (http:/tagc.univ-mrs.fr/pub/Cancer/).

Hybridization was done in excess of target (~15 ng of DNA in each spot) and binding of cDNAs to the targets was linear and proportional to the quantity of cDNA in the probe.

Detection and quantification of cDNA array hybridization signals

Quantitative data were obtained using an imaging plate device. Hybridization signal detection with a FUJI BAS 1500 machine and quantification with the HDG Analyzer

software (Genomic Solutions, Ann Arbor, MI) were done as previously described (http:/tagc.univ-mrs.fr/pub/Cancer/). Quantification was done by integrating all spot pixel intensities and substracting a spot background value determined in the neighboring area. Spots were located with a LaPlacian transformation. Spot background level was the median intensity of all the pixels present in a small window centered on the spot and which were not part of any spot (44). Quantified data were normalized in three steps and expressed as absolute gene expression levels (i.e. in percentage of abundance of individual mRNA with respect to mRNA within the sample), as described (4).

Array data analysis

Before analysis of the results, the reproducibility of the experiments was verified by comparing duplicate spots, or one hybridization with the same probe on two independent arrays, or two independent hybridizations with probes prepared from the same RNA. In every case, the results showed good reproducibility with respective correlation coefficients of 0.95, 0.98 and 0.98 (data not shown). Moreover, genes represented by two different clones on the array, such as CDK4 or ETV5, displayed similar expression profiles for the two clones in all samples. This reproducibility was sufficient enough to consider a 2-fold expression difference as significantly differential.

For graphical representation, data were displayed as absolute expression levels (Fig. 2a). For better visualization of clustering, results were log-transformed and displayed as relative values median-centered in each row and in each column (Fig. 2b). Hierarchical clustering was applied to the tissue samples and the genes using the Cluster program developed by Eisen (45) (average linkage clustering

using Pearson correlation as similarity metric). Results in Figs. 2 and 3 were displayed with the TreeView program (45).

Subsequent analysis was done using Excel software (Microsoft) and statistical analyses with the SPSS software. Metastasis-free survival and overall survival were measured from diagnosis until the first metastatic relapse or death respectively. They were estimated with the Kaplan-Meier method and compared between groups with the Log-Rank test. Correlations of gene pairs based on expression profiles were measured with the correlation coefficient r. The search for genes with expression levels correlated with tumor parameters was done in several successive steps.

First, genes were detected by comparing their median expression level in the two subgroups of tumors discordant according to the parameter of interest. The median values rather than the mean values were used because of the high variability of the expression levels for many genes, resulting in a standard deviation of expression level similar or superior to the mean value and making comparisons with means impossible. Second, these detected genes were inspected visually on graphics, and finally, an appropriate statistical analysis was applied to those that were convincing to validate the correlation. Comparison of GATA3 expression between ER-positive tumors and ER-negative tumors was validated using a Mann-Witney test. Correlation coefficients were used to compare the gene expression levels to the number of axillary nodes involved.

Northern blot analysis

Seventy-nine breast tumors, including 22 of the 34 tested on the arrays, were analyzed for GATA3 expression by Northern blot hybridization. RNA extraction from tumor samples and Northern blots were done as previously described

(43). The GATA3 probe was prepared from the IMAGE cDNA clone 129757, which corresponds to the 3' region (from +843 to +1689) of the GATA3 cDNA sequence (GenBank accession no. X55122). The insert (846 bp) was obtained by digestion of the clone with EcoRI and PacI enzymes. Northern blots were stripped and re-hybridized using a â-actin probe (46).

Fig. 1 shows an example of differential gene expression between normal breast tissue (NB) and breast tumor samples. Each cDNA array on Nylon filter was hybridized with a complex probe made from 5 μg of total RNA. The top image corresponds to the whole membrane. For the two bottom images, only the right portion of the membranes is shown. Numbers below the spots indicate housekeeping genes (1, GAPDH and 2, actin), negative control clones (3, 4 and 5) and examples of genes differentially expressed between NB and breast tumor (6, stromelysin3; 7, ERBB2; 8, MYBL2; 9, FOS; 10, TGFâR3; 11, desmin), and between ER- breast tumor and ER+ breast tumor (12, GATA3).

Fig. 2 is a representation of expression levels of 176 genes in normal breast tissue (NB) and 34 samples of breast carcinoma. Each column corresponds to a single tissue, and each row to a single gene. (a) The results are expressed as percentage abundance of individual mRNA within the sample, and are represented using a blue color scale. The color scale (log scale with a 3-fold interval) indicated at the bottom left ranges from light blue (expression level 0.001%) to dark blue (expression level > 3%). White squares indicate clones with undetectable expression levels and gray squares indicate missing data. The tissue samples are arbitrarily ordered and the clones are ordered from top to bottom according to increasing median expression levels. Horizontal black arrows on the right of the figure mark three clones with highly variable expression levels between the

tumors (stromelysin3, IGF2, GATA3 from top to bottom). (b) The results are shown as relative expression levels (relative to the median value of each row and each column) and are represented with a color scale indicated at the bottom left ranging from 1/100 to 100 fold changes (gray squares: missing data). Eighteen clones with median expression level equal to zero in the 34 tumors are omitted. The clustering program arranges samples (n = 35) along the horizontal axis so that those with the most similar expression profiles are placed adjacent to each other. Similarly, clones (n = 162) are near each other along the vertical axis if they show a strong expression profile correlation across all tissues. The length of the branches of the dendrograms capturing respectively the samples (top) and the clones (left) reflects the similarity of the related elements. Two groups of tumors are separated and color coded: group A (blue) and group B (orange). Horizontal black and horizontal red arrows on the right of the figure respectively mark three genes with highly variable expression levels between the tumors (IGF2, GATA3, stromelysin3 from top to bottom) and four pairs of different clones representing four genes. (c) Zoom representation of group A from Figure 2b, excluding the two outlyer tumors at the right. The clustering separates two subgroups of tumors, A1 and A2. The dotted branches correspond to tumors associated with metastatic relapse and death. Follow-up was longer in A2 than in A1 (median 81 months vs 47 for A1).

Fig. 3 is prognostic classification of breast cancer by gene expression profiling showing that gene expression-based tumour classification correlates with clinical outcome. The 12 samples of group A (see figure 2b and 2c) were reclustered using the top 32 differentially expressed genes between A1 and A2 subgroups. Data were displayed as in Fig. 2b and shown with the same color key.

The hierarchical clustering was applied to expression data from the 23 clones, out of 32, of which expression levels presented an at least two-fold change in at least two samples (out of 12). Two subgroups of tumors A1 and A2 are shown as well as two groups of differentially expressed clones. The dotted branches of tumor cluster A1 correspond to samples associated with metastatic relapse and death. Figure 3a shows Two-dimensional representation of hierarchical clustering results shown in figures 2a and 2b. The analysis delineates 4 groups of tumours A, B, C and D. Black squares indicate patients alive at last follow-up visit and red squares indicate patients who died. Three classes of patients with a statistically different clinical outcome were defined according to gene expression profiles: class A (n = 16), class B+C (n = 34), class D (n = 5). Figure 3b illustrates Kaplan-Meier plot of overall survival of the 3 classes of patients ($p<0.005$, log-rank test). And figure 3c illustrates Kaplan-Meier plot of metastasis-free survival of the 3 classes of patients ($p<0.05$, log-rank test).

Fig. 4 shows the correlation of GATA3 expression with ER phenotype. (a) The expression levels of GATA3 in 34 breast cancer samples (y axis) monitored by cDNA array analysis are reported in percentage of abundance of individual mRNA with respect to mRNA within the sample (log scale). GATA3 is significantly overexpressed in the ER-positive tumors (n = 23) versus the ER-negative tumors (n = 11) using the Mann-Witney test ($p = 0.0004$). The expression level of GATA3 in normal breast tissue is reported on the right (NB). (b) Northern blot analysis of GATA3 in normal breast sample (NB) and 9 breast cancer samples (AT: tumor analyzed with cDNA array and Northern blot; NT: tumor analyzed with Northern blot). Blots were probed successively

with cDNA from GATA3 (top) and â-actin (bottom). ER status is indicated for each tumor sample.

Data representation

Fig. 1 shows examples of hybridizations of cDNA arrays with probes made from RNA extracted from normal breast tissue and breast tumors.

The crude results of all hybridizations were processed to be presented either as absolute or relative values in schematic figures. The normalization procedure allowed display of absolute values expressed in percent of abundance of mRNA in the probe as shown in Fig. 2a. Each level of the blue color ladder represents a 3-fold interval of absolute abundance of mRNA. Each column corresponds to a tissue sample and each row to a gene. For graphic purposes, genes were ordered from top to bottom according to increasing median expression levels. Tumor samples were not ordered. The values in each sample displayed a wide range of intensities (3 decades in log scale) corresponding to expression levels ranging from approximately 0.002% to 5% of mRNA abundance. Many genes (see for example stromelysin 3, IGF2 and GATA3, arrows) displayed highly variable expression levels across all tumor samples, scattered over the whole dynamic range of values. A representation of relative values is shown in Fig. 2b. Absolute values were log-transformed, omitting 18 clones whose median intensity was equal to zero across all tissues. Data for each of the 162 remaining clones were then median-centered, as well as data for each sample, so that the relative variation was shown, rather than the absolute intensity. A color scale was used to display data: red for expression level higher than the median and green for expression level lower than the median. The magnitude of the deviation from the median was represented by

the color intensity. A hierarchical clustering program was then applied to group the 35 samples according to their overall gene expression profiles, and to group the 162 clones on the basis of similarity of their expression levels in all tissues. This resulted in a picture highlighting groups of correlated tissues and groups of correlated genes as depicted by dendrograms.

Breast tumor classification

As shown in Fig. 2b, the clustering algorithm identified two groups of samples, designated A (n = 15, including normal breast, NB) and B (n = 20). These groups were similar with respect to patient age, menopausal status at diagnosis, SBR grading and tumor pathological size. However, 72% of tumors in group A were node-positive and 75% in group B were node-negative. Moreover, 80% of the tumors in group B were estrogen receptor (ER) positive and 50% in group A were ER-negative. With a median follow-up of 44 months after diagnosis, overall survival was different between A and B groups: 5 women died in A (median follow-up 58 months) and 1 in B (median follow-up 40 months). But the frequency of metastatic relapse was relatively similar in the two groups, with 5 women who relapsed in A and 6 in B. Because the time between the diagnosis of metastasis and last follow-up is too short in B, a longer follow-up is needed to determine if these two different groups, defined with expression profiles, have really a different outcome with respect to overall survival.

In the group A of 15 samples, three samples (normal breast and two tumors) were different from each other and from the other 12 samples. The latter constituted two subgroups of tumors, A1 (n = 6) and A2 (n = 6), which could be further separated by clustering as shown in Fig. 2c. The

12 tumors had an uniformly high risk of metastatic relapse according to conventional prognostic features as shown in Table 1. Most of them had received comparable adjuvant anthracyclin-based chemotherapy after surgery, with more women treated in the A1 subgroup. Interestingly, these two subgroups, which could not be distinguished with commonly used histoclinical features, had a very different clinical outcome: there were 4 metastatic relapses and 4 deaths in A1 (median follow-up: 44 months). In contrast and despite a longer median follow-up (90 months), no metastasis or death occurred in A2. This resulted in a significant better metastasis-free survival (p 0.01) and overall survival (p 0.005) for group A2 than for group A1 tumors. No such subgrouping could be done in B.

TABLE 1

Subgroup	A1						A2					
Tumor position in the cluster	1	2	3	4	5	6	7	8	9	10	11	12
Age, years	46	58	60	63	51	58	46	47	50	47	46	66
Nodal status	1	0	0	16	13	37	10	4	1	2	0	0
Histological size, mm	60	20	26	35	20	30	27	25	30	25	20	22
SBR grade	II	III	II	III	II	III	II	II	II	II	II	III
ER status	neg	neg	neg	neg	neg	neg	pos	neg	pos	pos	pos	pos
Adjuvant chemotherapy	yes	yes	no	yes	yes	yes	yes	yes	no	yes	no	no
Metastasis	yes	no	yes	yes	no	yes	no	no	no	no	no	no
Follow-up, months	58	106	35	47	41	31	85	98	95	49	19	141
Patient status	D	A	D	D	A	D	A	A	A	A	A	A

Patient characteristics in subgroups A1 and A2. The 12 tumors are numbered from 1 to 12 according to their position from left to right in the clustering graphic displayed in Fig. 3. Adjuvant chemotherapy was anthracyclin-based. In the line concerning the patient status, A means alive and D means death from cancer progression.

Genes responsible for group A substructure were searched. These are potentially relevant to the prognosis and the sensitivity to chemotherapy in these tumors. Thirty-two genes out of 188 were identified by comparing their median expression level in A1 vs A2. Then, the 12 tumors were reclustered using the expression profiles of these genes as shown in Fig. 3. The same subgroups A1 and A2 were evident and separated by 2 groups of genes: as expected, high expression of ERBB2, MYC and EGFR was associated with bad prognosis subgroup A1 (6-8), and that of E-cadherin and the proto-oncogene MYB with good prognosis subgroup A2 (9, 10). For most of the other genes, these results may stimulate new investigations. Differentiation state is a good prognostic factor in breast cancer and, accordingly, genes associated with cell differentiation, such as GATA3 (11) and CRABP2 (12), had a high level of expression in the better outcome group. The high expression of Ephrin-A1 mRNA in the bad prognosis subgroup suggests a role of this growth factor in breast cancer and can be paralleled with its up-regulation during melanoma progression (13).

Differential gene expression between normal breast and breast tumors

To identify genes differentially expressed between breast tumors (T) and normal breast (NB), the NB value for each gene was compared to its expression level in each tumor. When the expression level of a gene in NB was undetectable, only qualitative information could be deduced and the mRNA was considered as differentially expressed if the signal intensity in the tumor was superior to the reproducibility threshold (0.002% of mRNA abundance). In the other cases, differential expression was defined by an at least 2-fold expression difference. Also, the number of

tumors where it was over- or underexpressed was measured. Table 2 shows a list of the top 20 over- and underexpressed genes. For these genes, the T/NB ratio is reported, where T represented their median expression value in the 34 tumors. This ratio ranged from 2.70 (ABCC5) to 17.76 (GATA3) for the overexpressed genes, and from 0.00 (desmin) to 0.29 (APC) for the underexpressed genes.

TABLE 2

Clone ID	Gene/Protein identity	Gene symbol	Chrom. location	N	T/NB
	Overexpressed genes				
154343	Granzyme H	GZMH	14q11.2	32	9,51
235947	Stromelysin 3	STMY3	22q11.2	31	15,92
207378	MYB Related Protein B	MYBL2	20q13.1	31	(a)
153275	Cellular Retinoic Acid Binding Protein 2	CRABP2	1q21.3	29	7,16
129757	GATA-binding protein 3	GATA3	10p15	28	17,76
120649	T-Lymphocyte surface CD2 antigen	CD2	1p13.1	28	7,54
109677	CREB Binding Protein	CREBBP	16p13.3	28	5,08
172152	EGFR-binding protein GRB2	GRB2	17q24-q25	28	5,00
66969	Transcription factor RELB	RELB	19	28	3,61
182007	ETS-Related Transcription Factor ELF1	ELF1	13q13	27	3,58
153446	LIM domain protein RIL	RIL	5q31.1	26	4,03
203394	ETS Variant gene 5 (ETS-related molecule)	ETV5	3q28	25	3,67
160963	Thrombospondin 1	THBS1	15q15	25	3,39
188393	POU domain, class 2, transcription Factor 2	POU2F2	19	24	4,02

Clone ID	Gene/Protein identity	Gene symbol	Chrom. location	N	T/NB
187822	Integrin, beta 2	ITGB2	21q22.3	24	3,01
243907	Nuclear Factor of Activating T cell Subunit p45	NF45	1	24	2,84
158347	EST H27202	EST		23	2,91
230933	EST AW184517	EST		22	2,85
212366	ATP-Binding Cassette, sub-family C (CFTR/MRP), 5	ABCC5	3q27	22	2,70
149401	Cathepsin D	CTSD	11p15.5	21	2,97
	Underexpressed genes				
153854	Desmin	DES	2q35	34	0,00
208717	P55-C-FOS proto-oncogene protein	FOS	14q24.3	33	0,05
159093	Transcription Factor AP4	TFAP4	16p13	33	0,11
124340	Tenascin XA	TNXA	6p21.3	33	0,14
133738	Prolactin	PRL	6p22.2-p21.3	32	0,00
133891	Chorionic Somatomammotropin Hormone 1	CSH1	17q22-q24	32	0,00
151501	Tyrosine Kinase Receptor TEK	TEK	9p21	32	0,00
183030	Activating Transcription Factor 3	ATF3	1	32	0,07
120916	Phosphodiesterase I	PDNP2	8q24.1	32	0,14
155716	EST R72075	EST		31	0,00
208118	Transforming Growth Factor Beta Receptor Type III	TGFBR3	1p33-p32	31	0,14
187547	Diphtheria Toxin Receptor	DTR	5q23	31	0,17
108490	HIV-1 Rev Binding protein	HRB	2q36	31	0,20
147002	B-cell CLL/lymphoma 2	BCL2	18q21.3	31	0,26
182610	Microsomal Glutathione S Transferase 1	MGST1	12p12.3-p12.1	31	0,28
152802	Phospholipase A2 Membrane	PLA2G2A	1p35	30	0,03

Clone ID	Gene/Protein identity	Gene symbol	Chrom. location	N	T/NB
	Associated, group IIA				
183087	Interleukin 3 Receptor Alpha chain	IL3RA	Xp22.3;Yp13.3	30	0,24
108571	Retinoblastoma-Like 2 (p130)	RBL2	16q12.2	29	0,28
125294	Adenomatous Polyposis Coli Protein	APC	5q21-q22	29	0,29
151767	FASL Receptor	TNFRSF6	10q24.1	28	0,27

List of the genes that show the most frequent differential expression between normal breast tissue and 34 breast carcinomas as measured by cDNA array analysis. N indicates the number of tumor samples where the gene is dysregulated (fold change > 2) compared to normal breast tissue. T/NB represents the ratio: median expression level in 34 breast tumors / expression level in normal breast. (a) MYBL2 transcript displayed a median expression level of 0.025% in breast tumors and was undetectable in NB.

High expression of mucin 1, NM23, ERBB2, FGFR1 and FGFR2, MYC, stromelysin3, cathepsin D and downregulation of FOS, APC, RBL2, FAS, BCL2 were found, reflecting what is known about their biology in cancer. GATA3, which codes for a member of the GATA family of zinc finger transcription factors, and CRABP2, encoding one of the two cellular retinoic acid-binding proteins, showed high expression of mRNA, extending previous results on cDNA arrays (4).

Differential gene expression among various breast tumors and correlation with histoclinical prognostic parameters

To search for potential prognostic markers in breast cancer, genes with expression levels correlated with conventional histoclinical prognostic parameters were looked for: age of patients, axillary node status, tumor size, histological grade and ER status. No significant correlation was found with age, tumor size and histological grade. However, the expression profiles of some genes correlated with ER status and axillary node involvement.

To identify genes potentially relevant to the hormone-responsive phenotype, the gene expression profiles in ER-positive breast cancers (n = 23) vs ER-negative breast cancers (n = 11) were compared. Sixteen clones displayed a median intensity of 0 in both groups. Twenty-five presented a fold change superior to 2. Table 3a displays the top 10 over- and underexpressed genes. Among them, the most differentially expressed was GATA3 with a median intensity ratio ER+/ER- of 28.6 and a value for the first quartile of ER-positive tumors superior (5-fold) to the value of the third quartile of the ER-negative tumors as shown in Fig. 4a. The high expression of GATA3 in ER-positive tumors was statistically significant using a Mann-Witney test ($p < 0.001$). All ER-positive tumors and only 18% of ER-negative tumors displayed a GATA3 expression level greatly superior (fold change > 3) to the normal breast value. Furthermore GATA3 expression was analyzed by Northern blot hybridization (Fig. 4b) in a panel of 79 breast cancers (21 ER-negative tumors and 58 ER-positive tumors), including 22 of the tumors analyzed with cDNA arrays. It confirmed the array results for those 22 tumors as well as the strong correlation between ER status and GATA3 RNA expression (Mann-Witney test, $p \leq 0.0001$).

TABLE 3A

Clone ID	Gene/Protein identity	Gene symbol	ER+/ER-
129757	GATA-binding protein 3	GATA3	28,6
356763	Granzyme A	GZMA	5,7
248613	MYB proto-oncogene	MYB	3,4
211999	KIAA1075 protein	KIAA1075	3,3
235947	Stromelysin 3	STMY3	3,1
229839	Macrophage Stimulating 1	MST1	2,8
153275	Cellular Retinoic Acid Binding Protein 2	CRABP2	2,7
301950	X-box Binding Protein 1	XBP1	2,7
205314	Tumor Protein p53	TP53	2,5
126233	Insulin-like Growth Factor 2	IGF2	2,4
66322	CD3G antigen, Gamma	CD3G	0,0
195022	Interleukin 2 Receptor Gamma chain	IL2RG	0,0
111461	SOX4 Protein	SOX4	0,4
151475	Epidermal Growth Factor Receptor	EGFR	0,5
195022	Interleukin 2 Receptor Beta chain	IL2RB	0,5
130788	Topoisomerase (DNA) II beta (180kD)	TOP2B	0,6
323948	SOX9 Protein	SOX9	0,6
183641	S100 calcium-binding protein Beta	S100B	0,6
246620	EST N53133	EST	0,6
231424	Glutathione S Transferase Pi	GSTP1	0,6

To search for genes whose expression profile was correlated with axillary lymph node status, a strong prognostic factor in breast cancer, the group of node-negative tumors (n = 19) was compared with the group of tumors with massive axillary extension (10 or more positive nodes). Furthermore, because survival decreases with the increase of the number of tumor-involved lymph nodes and because the expression measurements were quantitative, it was looked for a correlation between the expression levels of

these genes and the number of tumor-involved nodes (quantitative variables). Table 3b shows a list of the top 10 over- and underexpressed genes between these 2 groups. Most of these genes have not been previously reported as associated with node status, but some of these results are in agreement with literature data. The gene encoding the tyrosine kinase receptor ERBB2 was the most significantly overexpressed gene in node-positive tumors and displayed the highest correlation coefficient ($r = 0.68$; $p \leq 0.0001$).

TABLE 3B

Clone ID	Gene/Protein identity	Gene symbol	N-/10N+
129757	GATA-binding protein 3	GATA3	11,0
160963	Thrombospondin 1	THBS1	6,6
151475	Epidermal Growth Factor Receptor	EGFR	5,4
120916	Phosphodiesterase I	PDNP2	4,9
183030	Activating Transcription Factor 3	ATF3	4,6
211999	KIAA1075 protein	KIAA1075	4,5
110480	Nuclear Factor 1 A-type	NF1A	4,5
182264	P-Selectin	SELP	4,4
356763	Granzyme A	GZMA	4,3
214008	E-cadherin	CDH1	4,0
147016	ERBB2 Receptor Protein-Tyrosine Kinase	ERBB2	0,2
179197	Protein Phosphatase PP2A, 55 kD Subunit	PP2A BR gamma	0,2
231424	Glutathione S Transferase Pi	GSTP1	0,4
111461	SOX4 Protein	SOX4	0,4
195022	Interleukin 2 Receptor Beta chain	IL2RB	0,4
220451	Zinc Finger protein 144	ZNF144	0,5
125413	Mucin 1	MUC1	0,6
290007	CD44 antigen, epithelial form	CD44	0,6
108571	Retinoblastoma-Like 2 (p130)	RBL2	0,7
130788	Topoisomerase (DNA) II Beta (180kD)	TOP2B	0,7

Gene clusters

Gene clustering from Fig. 2b showed groups of genes with correlated expression across samples. When different clones represented the same gene, they were

clustered next to each other (red arrows). Correlation coefficients between gene pairs in the 34 tumors were often high (1% of the 13,041 gene pairs showed a correlation coefficient superior to 0.95 - not shown). An example of highly correlated gene expression is that of BCL2 and RBL2. Such correlated expression, although it has not been described in the literature, probably reflects a common mechanism of regulation for these two genes. Furthermore, these genes also exhibited significant correlated expression with other genes such as PPP2CA, AKT2, PRKCSH or TNFRSF6/FAS. In particular, a striking correlated expression between BCL2 and FAS could be observed (r = 0.91; data not shown). The exact meaning of this correlation is unknown, although it may reflect the necessary balance between apoptosis and anti-apoptosis for cell survival.

Although in human cancer the proportion of changes that is reflected at the RNA level is not known, monitoring gene expression patterns appears as a very promising way of increasing the knowledge of the disease. Several different types of cancer have been investigated using cDNA arrays: cervical (14), hepatocellular (15), ovarian (16), colon (17) and renal carcinomas (18), glioblastomas (19), melanomas (20) (21), rhabdomyosarcomas (22), acute leukemias (23) and lymphomas (24). In breast cancer, pioneering studies have yielded the first expression patterns (4, 25-31). They have in particular addressed the important issue of molecular differences in hormone responsive and non-responsive breast tumors. Thus, Yang et al. (28) and Hoch et al. (25) compared expression profiles of breast carcinoma cell lines known to represent these two categories and identified a few genes with differential expression. One of these genes was GATA3. In these studies, cell lines were mostly used and tumor samples were rarely

tested and generally in small numbers. The first study analyzing the expression profiles of a large series of breast cancers was published recently (32), but no correlation with clinical outcome was mentioned.

Several interesting points can be made based on the present experimentation. First, the differences in expression patterns among the tumors provided molecular transcriptional evidence of the histoclinical heterogeneity of breast cancer. This diversity was multifactorial, linked to many different genes, highlighting the interest of high throughput analysis in this context. It was possible, with a hierarchical clustering program integrating the expression profiles, to separate normal breast tissue from most tumors and, moreover, to identify two different groups of tumors. Most importantly, two different subgroups of tumors with a very distinct clinical outcome that could not be predicted with classical prognostic factors have been identified by clustering. Indeed, all these tumors had a theoretically bad prognosis as evaluated by current histoclinical tools. All these patients would be at the present time treated with adjuvant chemotherapy, but without the capacity for the physicians to identify patients who will benefit of this treatment and those who will not benefit.

Gene expression profiles were able to make this discrimination. Such predictive tools have important therapeutic implications. Patients with features of poor prognosis are candidates for other treatment than standard chemotherapy, avoiding loss of time and toxicities related to first-line chemotherapy. These results suggest that the histoclinical category of poor prognosis breast cancer, currently treated with adjuvant anthracyclin-based chemotherapy, groups together at least two molecularly distinct subgroups of tumors with different outcome which

would require distinct chemotherapy regimens. Expression profiles could thus provide a new and more accurate way of classifying breast tumors of poor prognosis and managing patients.

Similarly, despite molecular heterogeneity, significant correlations between the expression level of genes (GATA3, ERBB2) and histological tumor parameters were identified. The ER-positivity in breast cancer has been correlated with tumor differentiation, low proliferating rate, favorable prognosis and response to hormonal therapy. The relation between hormone sensitivity of breast cancer and ER status is not perfect, and it is possible that some genes related to ER expression are more important than ER to characterize the hormone sensitive phenotype. These genes could serve as predictive factors to guide the therapy.

GATA3 mRNA expression was highly correlated with ER status. GATA3, which is not estrogen-regulated (25), is a transcription factor that could regulate the expression of genes involved in the ER-positive phenotype. Among the other genes that were found associated with ER status during the experimental work leading to the present invention, some, such as MYB (10), stromelysin 3 (33), and CRABP2 (34), have been previously reported expressed at high levels in ER-positive breast tumors. The higher levels of TP53 mRNA in ER-positive tumors studied were surprising, although in agreement with a recent study (27). Most studies concerning TP53 expression analyzed the protein level rather than the mRNA level, and TP53 protein levels are classically negatively correlated with the ER status (35). The high expression of CRABP2 could be related to the better differentiated status of the ER-positive tumors. The low expression of the three immunity-related genes IL2RB, IL2RG and CD3G may be related to the low lymphoid infiltration in

these well differentiated tumors. ERBB2 high expression in breast cancer has been associated with a poor prognosis and some resistance to hormonal therapy and chemotherapy (36). It is involved in the regulation of cellular differentiation, adhesion, and motility. The motility-enhancing activity of ERBB2 (37) could be responsible for the increased metastatic potential and the unfavorable prognosis of the breast tumors that overexpress ERBB2. The low expression of E-cadherin and thrombospondin 1 in node-positive tumors are consistent with their putative role in different steps of metastatic spread: E-cadherin is an epithelial cell adhesion molecule whose disturbance is a prerequisite for the release of invasive cells in carcinomas (38) and thrombospondin 1 inhibits angiogenesis (39). Similarly, the high expression of the molecule surface antigen Mucin 1 in node-positive tumors (40) can reduce cell-cell interactions facilitating cell detachment and metastasis. CD44, encoding a transmembrane glycoprotein involved in cell adhesion and lymph node homing (41) was expressed at high levels in node-positive tumors as well as GSTP1 (Glutathione-S-Transferase Pi), recently reported associated with increased tumor size (27).

Second, there were a number of genes with highly correlated expression patterns. Gene correlations have already been reported with larger series of genes, essentially under dynamic experimental conditions (42) and recently in steady states (17). Here, correlations were based on expression profiles of a relatively small but selected series of genes and in steady states represented by different breast tumors. Gene correlations are potentially useful tools for cancer research in two ways: i)- they can provide information about the general regulation circuitry of a cancerous cell, allowing the identification of regulatory elements controlling expression networks; ii)- they offer the

possibility of reducing the complexity of the system analyzed by replacing, for example, the intensities of a large number of genes present in a gene cluster by their respective mean intensities.

Finally, these results highlight the great potential of cDNA array in cancer research. The gene expression profiles confirmed the heterogeneity of breast cancer, and most importantly allowed us to identify, among a series of poor prognosis breast tumors, two subtypes of the disease not yet recognized with usual histoclinical parameters but with a different clinical outcome after adjuvant chemotherapy. Furthermore, the present invention allows detecting genes of which expression was correlated with classical prognostic factors.

Table 4 displays a library of polynucleotides SEQ ID NO :1 to SEQ ID NO : 468 corresponding to a population of polynucleotide sequences underexpressed or overexpressed in cells derived from tumors, more particularly breast tumors, and their respective complements.

TABLE 4

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
HRB	1	hiv-1 rev binding protein	SEQ ID No:1		SEQ ID No:2
GATA1	2	gata-binding protein 1 (globin transcription factor 1)		SEQ ID No:3	SEQ ID No:4
TLK2	3	tousled-like kinase 2		SEQ ID No:5	SEQ ID No:6
EST T81919	4	ests, weakly similar to alu7_human alu subfamily sq sequence contamination warning entry [h.sapiens]	SEQ ID No:7	SEQ ID No:8	
CCND1	5	cyclin d1 (prad1: parathyroid adenomatosis 1)	SEQ ID No:9		SEQ ID No:10
STAT1	6	signal transducer and activator of transcription 1, 91kd		SEQ ID No:11	SEQ ID No:12
FGFR2	7	fibroblast growth factor receptor 2 (bacteria-expressed kinase, keratinocyte growth factor receptor, craniofacial dysostosis 1, crouzon syndrome, pfeiffer syndrome, jackson-weiss syndrome)	SEQ ID No:13	SEQ ID No:14	SEQ ID No:15
EST T89980	8	ests	SEQ ID No:16		
PPP3CC	9	protein phosphatase 3 (formerly 2b), catalytic subunit, gamma isoform (calcineurin a gamma)	SEQ ID No:17	SEQ ID No:18	SEQ ID No:19
EST T90726	10	ests	SEQ ID No:20	SEQ ID No:21	
SOX4	11	sry (sex determining region y)-box 4	SEQ ID No:22	SEQ ID No:23	SEQ ID No:24
RNF5	12	ring finger protein 5		SEQ ID No:25	SEQ ID No:26
AXL	13	axl receptor tyrosine kinase	SEQ ID No:27	SEQ ID No:28	SEQ ID No:29
CTSB	14	cathepsin b		SEQ ID No:30	SEQ ID No:31
PPP4C	15	protein phosphatase 4 (formerly x), catalytic subunit	SEQ ID No:32	SEQ ID No:33	SEQ ID No:34
EST T79867	16	ests	SEQ ID No:35		
FGFR4	17	fibroblast growth factor receptor 4	SEQ ID No:36	SEQ ID No:37	SEQ ID No:38
ENPP2	18	ectonucleotide pyrophosphatase/phosphodiesterase 2	SEQ ID No:39	SEQ ID No:40	SEQ ID No:41

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
		(autotaxin)			
RELA	19	v-rel avian reticuloendotheliosis viral oncogene homolog a (nuclear factor of kappa light polypeptide gene enhancer in b-cells 3 (p65))	SEQ ID No:42		SEQ ID No:43
ITK	20	il2-inducible t-cell kinase		SEQ ID No:44	SEQ ID No:45
TNXB	21	tenascin xb		SEQ ID No:46	SEQ ID No:47
CSF1	22	colony stimulating factor 1 (macrophage)	SEQ ID No:48	SEQ ID No:49	SEQ ID No:50
VIL2	23	villin 2 (ezrin)	SEQ ID No:51	SEQ ID No:52	SEQ ID No:53
APC	24	adenomatosis polyposis coli	SEQ ID No:54	SEQ ID No:55	SEQ ID No:56
MUC1	25	mucin 1, transmembrane		SEQ ID No:57	SEQ ID No:58
IGF2	26	insulin-like growth factor 2 (somatomedin a)	SEQ ID No:59	SEQ ID No:60	SEQ ID No:61
EMR1	27	egf-like module containing, mucin-like, hormone receptor-like sequence 1	SEQ ID No:62	SEQ ID No:63	SEQ ID No:64
KIAA0427	28	kiaa0427 gene product	SEQ ID No:65	SEQ ID No:66	SEQ ID No:67
SYK	29	spleen tyrosine kinase	SEQ ID No:68	SEQ ID No:69	SEQ ID No:70
IL7R	30	interleukin 7 receptor		SEQ ID No:71	SEQ ID No:72
MYC	31	v-myc avian myelocytomatosis viral oncogene homolog	SEQ ID No:73	SEQ ID No:74	SEQ ID No:75
GATA3	32	gata-binding protein 3	SEQ ID No:76	SEQ ID No:77	SEQ ID No:78
GRB7	33	growth factor receptor-bound protein 7	SEQ ID No:79	SEQ ID No:80	SEQ ID No:81
TOP2B	34	topoisomerase (dna) ii beta (180kd)		SEQ ID No:82	SEQ ID No:83
CASP4	35	caspase 4, apoptosis-related cysteine protease	SEQ ID No:84		SEQ ID No:85
TIMP2	36	tissue inhibitor of metalloproteinase 2		SEQ ID No:86	SEQ ID No:87
DDT	37	d-dopachrome tautomerase	SEQ ID No:88	SEQ ID No:89	SEQ ID No:90
PRL	38	prolactin	SEQ ID	SEQ ID	SEQ ID

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
			No:91	No:92	No:93
PRLR	39	prolactin receptor	SEQ ID No:94	SEQ ID No:95	SEQ ID No:96
IL2RB	40	interleukin 2 receptor, beta	SEQ ID No:97	SEQ ID No:98	SEQ ID No:99
GATA3	41	gata-binding protein 3	SEQ ID No:100	SEQ ID No:101	SEQ ID No:78
PGF	42	placental growth factor, vascular endothelial growth factor-related protein		SEQ ID No:102	SEQ ID No:103
UBE3A	43	ubiquitin protein ligase e3a (human papilloma virus e6-associated protein, angelman syndrome)		SEQ ID No:104	SEQ ID No:105
TC21	44	oncogene tc21	SEQ ID No:106	SEQ ID No:107	SEQ ID No:108
TIE	45	tyrosine kinase with immunoglobulin and epidermal growth factor homology domains		SEQ ID No:109	SEQ ID No:110
AMFR	46	autocrine motility factor receptor	SEQ ID No:111	SEQ ID No:112	SEQ ID No:113
EST R81127	47	homo sapiens mrna; cdna dkfzp434c136 (from clone dkfzp434c136)	SEQ ID No:114		
BCL2	48	b-cell cll/lymphoma 2	SEQ ID No:115	SEQ ID No:116	SEQ ID No:117
ERBB2	49	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2 (neuro/glioblastoma derived oncogene homolog)		SEQ ID No:118	SEQ ID No:119
MDM2	50	mouse double minute 2, human homolog of; p53-binding protein		SEQ ID No:120	SEQ ID No:121
GATA3	51	gata-binding protein 3	SEQ ID No:122		SEQ ID No:78
HIP-55	52	src homology 3 domain-containing protein hip-55	SEQ ID No:123	SEQ ID No:124	SEQ ID No:125
CTSD	53	cathepsin d (lysosomal aspartyl protease)	SEQ ID No:126	SEQ ID No:127	SEQ ID No:128
IGF1R	54	insulin-like growth factor 1 receptor		SEQ ID No:129	SEQ ID No:130
INSR	55	insulin receptor		SEQ ID No:131	SEQ ID No:132
FOXO1A	56	forkhead box o1a (rhabdomyosarcoma)		SEQ ID No:133	SEQ ID No:134
EGFR	57	epidermal growth factor receptor	SEQ ID	SEQ ID	SEQ ID

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
		(avian erythroblastic leukemia viral (v-erb-b) oncogene homolog)	No:135	No:136	No:137
TEK	58	tek tyrosine kinase, endothelial (venous malformations, multiple cutaneous and mucosal)	SEQ ID No:138	SEQ ID No:139	SEQ ID No:140
TNFRSF6	59	tumor necrosis factor receptor superfamily, member 6	SEQ ID No:141	SEQ ID No:142	SEQ ID No:143
CDKN1A	60	cyclin-dependent kinase inhibitor 1a (p21, cip1)	SEQ ID No:144	SEQ ID No:145	SEQ ID No:146
PLA2G2A	61	phospholipase a2, group iia (platelets, synovial fluid)	SEQ ID No:147	SEQ ID No:148	SEQ ID No:149
GAPD	62	glyceraldehyde-3-phosphate dehydrogenase	SEQ ID No:150	SEQ ID No:151	SEQ ID No:152
JUNB	63	jun b proto-oncogene	SEQ ID No:153	SEQ ID No:154	SEQ ID No:155
CRABP2	64	cellular retinoic acid-binding protein 2	SEQ ID No:156	SEQ ID No:157	SEQ ID No:158
ACVRL1	65	activin a receptor type ii-like 1	SEQ ID No:159	SEQ ID No:160	SEQ ID No:161
RIL	66	lim domain protein		SEQ ID No:162	SEQ ID No:163
SHC1	67	shc (src homology 2 domain-containing) transforming protein 1		SEQ ID No:164	SEQ ID No:165
GAPD	68	glyceraldehyde-3-phosphate dehydrogenase	SEQ ID No:166	SEQ ID No:167	SEQ ID No:152
DES	69	desmin	SEQ ID No:168	SEQ ID No:169	SEQ ID No:170
CSNK2B	70	casein kinase 2, beta polypeptide		SEQ ID No:171	SEQ ID No:172
GLG1	71	golgi apparatus protein 1	SEQ ID No:173	SEQ ID No:174	SEQ ID No:175
EDNRB	72	endothelin receptor type b		SEQ ID No:176	SEQ ID No:177
GZMB	73	granzyme b (granzyme 2, cytotoxic t-lymphocyte-associated serine esterase 1)	SEQ ID No:178		SEQ ID No:179
FGFR1	74	fibroblast growth factor receptor 1 (fms-related tyrosine kinase 2, pfeiffer syndrome)	SEQ ID No:180	SEQ ID No:181	SEQ ID No:182
PPP2CA	75	protein phosphatase 2 (formerly 2a), catalytic subunit, alpha isoform		SEQ ID No:183	SEQ ID No:184
EST R55460	76	homo sapiens, clone image:4054156, mRNA, partial cds		SEQ ID No:185	

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
IGKC	77	immunoglobulin kappa constant	SEQ ID No:186		
MC1R	78	melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)		SEQ ID No:187	SEQ ID No:188
NRG1	79	neuregulin 1	SEQ ID No:189	SEQ ID No:190	SEQ ID No:191
CNTFR	80	ciliary neurotrophic factor receptor		SEQ ID No:192	SEQ ID No:193
ANG	81	angiogenin, ribonuclease, rnase a family, 5		SEQ ID No:194	SEQ ID No:195
ENG	82	endoglin (osler-rendu-weber syndrome 1)	SEQ ID No:196	SEQ ID No:197	SEQ ID No:198
EGF	83	epidermal growth factor (beta-urogastrone)	SEQ ID No:199		SEQ ID No:200
HRMT1L1	84	hmt1 (hnrnp methyltransferase, s. cerevisiae)-like 1	SEQ ID No:201	SEQ ID No:202	SEQ ID No:203
ETV4	85	ets variant gene 4 (e1a enhancer-binding protein, e1af)	SEQ ID No:204	SEQ ID No:205	
ANXA11	86	annexin a11		SEQ ID No:206	SEQ ID No:207
PDGFRB	87	platelet-derived growth factor receptor, beta polypeptide		SEQ ID No:208	SEQ ID No:209
WBSCR14	88	williams-beuren syndrome chromosome region 14		SEQ ID No:210	SEQ ID No:211
CD74	89	cd74 antigen (invariant polypeptide of major histocompatibility complex, class ii antigen-associated)		SEQ ID No:212	SEQ ID No:213
ANXA7	90	annexin a7		SEQ ID No:214	SEQ ID No:215
THBS1	91	thrombospondin 1	SEQ ID No:216		SEQ ID No:217
PTPN2	92	protein tyrosine phosphatase, non-receptor type 2	SEQ ID No:218	SEQ ID No:219	SEQ ID No:220
EPHA2	93	epha2	SEQ ID No:221		SEQ ID No:222
TIMP1	94	tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor)	SEQ ID No:223	SEQ ID No:224	SEQ ID No:225
EFNA1	95	ephrin-a1		SEQ ID No:226	SEQ ID No:227

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
EDNRA	96	endothelin receptor type a	SEQ ID No:228		SEQ ID No:229
GRB2	97	growth factor receptor-bound protein 2	SEQ ID No:230	SEQ ID No:231	SEQ ID No:232
JUND	98	jun d proto-oncogene	SEQ ID No:233		SEQ ID No:234
SMARCA2	99	swi/snf related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2	SEQ ID No:235	SEQ ID No:236	SEQ ID No:237
PPP2R2C	100	protein phosphatase 2 (formerly 2a), regulatory subunit b (pr 52), gamma isoform	SEQ ID No:238	SEQ ID No:239	
THBS3	101	thrombospondin 3	SEQ ID No:240		SEQ ID No:241
ACTG1	102	actin, gamma 1	SEQ ID No:242	SEQ ID No:243	SEQ ID No:244
ITGA6	103	integrin, alpha 6	SEQ ID No:245	SEQ ID No:246	SEQ ID No:247
RAD9	104	rad9 (s. pombe) homolog	SEQ ID No:248		SEQ ID No:249
ATF3	105	activating transcription factor 3	SEQ ID No:250	SEQ ID No:251	SEQ ID No:252
AKT2	106	v-akt murine thymoma viral oncogene homolog 2	SEQ ID No:253		SEQ ID No:254
S100B	107	s100 calcium-binding protein, beta (neural)		SEQ ID No:255	SEQ ID No:256
ABCB1	108	atp-binding cassette, sub-family b (mdr/tap), member 1	SEQ ID No:257		SEQ ID No:258
SELE	109	selectin e (endothelial adhesion molecule 1)	SEQ ID No:259	SEQ ID No:260	SEQ ID No:261
EGF	110	epidermal growth factor (beta-urogastrone)	SEQ ID No:262		SEQ ID No:200
PRKCSH	111	protein kinase c substrate 80k-h		SEQ ID No:263	SEQ ID No:264
DTR	112	diphtheria toxin receptor (heparin-binding epidermal growth factor-like growth factor)		SEQ ID No:265	SEQ ID No:266
ITGB2	113	integrin, beta 2 (antigen cd18 (p95), lymphocyte function-associated antigen 1; macrophage antigen 1 (mac-1) beta subunit)		SEQ ID No:267	SEQ ID No:268
NEO1	114	neogenin (chicken) homolog 1		SEQ ID No:269	SEQ ID No:270
POU2F2	115	pou domain, class 2, transcription	SEQ ID		SEQ ID

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
		factor 2	No:271		No:272
BIRC4	116	baculoviral iap repeat-containing 4	SEQ ID No:273		SEQ ID No:274
DAP3	117	death associated protein 3	SEQ ID No:275		SEQ ID No:276
GNRH1	118	gonadotropin-releasing hormone 1 (leutinizing-releasing hormone)		SEQ ID No:277	SEQ ID No:278
IL2RG	119	interleukin 2 receptor, gamma (severe combined immunodeficiency)	SEQ ID No:279	SEQ ID No:280	SEQ ID No:281
DAP3	120	death associated protein 3	SEQ ID No:282	SEQ ID No:283	SEQ ID No:276
PTK2	121	ptk2 protein tyrosine kinase 2		SEQ ID No:284	SEQ ID No:285
CDK4	122	cyclin-dependent kinase 4	SEQ ID No:286	SEQ ID No:287	SEQ ID No:288
BTF3	123	basic transcription factor 3	SEQ ID No:289		SEQ ID No:290
CSF1R	124	colony stimulating factor 1 receptor, formerly mcdonough feline sarcoma viral (v-fms) oncogene homolog	SEQ ID No:291		SEQ ID No:292
FLI1	125	friend leukemia virus integration 1	SEQ ID No:293	SEQ ID No:294	SEQ ID No:295
EST R97218	126	ests, highly similar to tvhume hepatocyte growth factor receptor precursor [h.sapiens]	SEQ ID No:296	SEQ ID No:297	
ETV5	127	ets variant gene 5 (ets-related molecule)	SEQ ID No:298	SEQ ID No:299	SEQ ID No:300
CDK4	128	cyclin-dependent kinase 4	SEQ ID No:301	SEQ ID No:302	SEQ ID No:288
YES1	129	v-yes-1 yamaguchi sarcoma viral oncogene homolog 1	SEQ ID No:303		SEQ ID No:304
IFI75	130	interferon-induced protein 75, 52kd	SEQ ID No:305	SEQ ID No:306	SEQ ID No:307
MYBL2	131	v-myb avian myeloblastosis viral oncogene homolog-like 2	SEQ ID No:308	SEQ ID No:309	SEQ ID No:310
TGFBR3	132	transforming growth factor, beta receptor iii (betaglycan, 300kd)	SEQ ID No:311	SEQ ID No:312	SEQ ID No:313
PRDX2	133	peroxiredoxin 2	SEQ ID No:314	SEQ ID No:315	SEQ ID No:316
FOS	134	v-fos fbj murine osteosarcoma viral oncogene homolog		SEQ ID No:317	SEQ ID No:318

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
RBBP7	135	retinoblastoma-binding protein 7	SEQ ID No:319	SEQ ID No:320	SEQ ID No:321
KIAA1075	136	kiaa1075 protein	SEQ ID No:322	SEQ ID No:323	
ABCC5	137	atp-binding cassette, sub-family c (cftr/mrp), member 5		SEQ ID No:324	SEQ ID No:325
CDH1	138	cadherin 1, type 1, e-cadherin (epithelial)	SEQ ID No:326	SEQ ID No:327	SEQ ID No:328
ZNF144	139	zinc finger protein 144 (mel-18)		SEQ ID No:329	SEQ ID No:330
MST1	140	macrophage stimulating 1 (hepatocyte growth factor-like)	SEQ ID No:331	SEQ ID No:332	SEQ ID No:333
GSTP1	141	glutathione s-transferase pi	SEQ ID No:334	SEQ ID No:335	SEQ ID No:336
BCL2	142	b-cell cll/lymphoma 2	SEQ ID No:337	SEQ ID No:338	SEQ ID No:117
PCNA	143	proliferating cell nuclear antigen	SEQ ID No:339	SEQ ID No:340	SEQ ID No:341
BS69	144	adenovirus 5 e1a binding protein	SEQ ID No:342	SEQ ID No:343	SEQ ID No:344
MMP11	145	matrix metalloproteinase 11 (stromelysin 3)	SEQ ID No:345		SEQ ID No:346
MGC13071	146	hypothetical protein mgc13071	SEQ ID No:347	SEQ ID No:348	SEQ ID No:349
ILF2	147	interleukin enhancer binding factor 2, 45kd		SEQ ID No:350	SEQ ID No:351
FLJ11307	148	hypothetical protein flj11307	SEQ ID No:352		SEQ ID No:353
MYB	149	v-myb avian myeloblastosis viral oncogene homolog		SEQ ID No:354	SEQ ID No:355
ZNF9	150	zinc finger protein 9 (a cellular retroviral nucleic acid binding protein)	SEQ ID No:356		SEQ ID No:357
CREM	151	camp responsive element modulator	SEQ ID No:358	SEQ ID No:359	SEQ ID No:360
CTSB	152	cathepsin b	SEQ ID No:361		SEQ ID No:31
MLANA	153	melan-a	SEQ ID No:362	SEQ ID No:363	SEQ ID No:364
APR-1	154	apr-1 protein	SEQ ID No:365	SEQ ID No:366	SEQ ID No:367
ETV5	155	ets variant gene 5 (ets-related	SEQ ID	SEQ ID	SEQ ID

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
		molecule)	No:368	No:369	No:300
CD69	156	cd69 antigen (p60, early t-cell activation antigen)		SEQ ID No:370	SEQ ID No:371
TC21	157	oncogene tc21	SEQ ID No:372	SEQ ID No:373	SEQ ID No:108
CD44	158	cd44 antigen (homing function and indian blood group system)	SEQ ID No:374	SEQ ID No:375	SEQ ID No:376
CDKN3	159	cyclin-dependent kinase inhibitor 3 (cdk2-associated dual specificity phosphatase)	SEQ ID No:377	SEQ ID No:378	SEQ ID No:379
MXI1	160	max-interacting protein 1		SEQ ID No:380	SEQ ID No:381
HOXA5	161	homeo box a5	SEQ ID No:382	SEQ ID No:383	SEQ ID No:384
XBP1	162	x-box binding protein 1	SEQ ID No:385	SEQ ID No:386	SEQ ID No:387
TNFAIP3	163	tumor necrosis factor, alpha-induced protein 3	SEQ ID No:388	SEQ ID No:389	SEQ ID No:390
SRF	164	serum response factor (c-fos serum response element-binding transcription factor)	SEQ ID No:391	SEQ ID No:392	SEQ ID No:393
SOX9	165	sry (sex determining region y)-box 9 (campomelic dysplasia, autosomal sex-reversal)	SEQ ID No:394		SEQ ID No:395
CDH15	166	cadherin 15, m-cadherin (myotubule)	SEQ ID No:396	SEQ ID No:397	SEQ ID No:398
BCL2	167	b-cell cll/lymphoma 2	SEQ ID No:399	SEQ ID No:400	SEQ ID No:117
EST W73386	168	ests	SEQ ID No:401		
GZMA	169	granzyme a (granzyme 1, cytotoxic t-lymphocyte-associated serine esterase 3)	SEQ ID No:402		SEQ ID No:403
FOS	170	v-fos fbj murine osteosarcoma viral oncogene homolog	SEQ ID No:404	SEQ ID No:405	SEQ ID No:318
ILF1	171	interleukin enhancer binding factor 1	SEQ ID No:406	SEQ ID No:407	SEQ ID No:408
ARHGDIA	172	rho gdp dissociation inhibitor (gdi) alpha	SEQ ID No:409	SEQ ID No:410	SEQ ID No:411
C4A	173	complement component 4a	SEQ ID No:412		SEQ ID No:413
CD3G	174	cd3g antigen, gamma polypeptide (tit3 complex)	SEQ ID No:414	SEQ ID No:415	SEQ ID No:416

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
RELB	175	v-rel avian reticuloendotheliosis viral oncogene homolog b (nuclear factor of kappa light polypeptide gene enhancer in b-cells 3)	SEQ ID No:417	SEQ ID No:418	SEQ ID No:419
ESR1	176	estrogen receptor 1	SEQ ID No:420	SEQ ID No:421	SEQ ID No:422
PBX1	177	pre-b-cell leukemia transcription factor 1	SEQ ID No:423	SEQ ID No:424	SEQ ID No:425
GLI3	178	gli-kruppel family member gli3 (greig cephalopolysyndactyly syndrome)	SEQ ID No:426	SEQ ID No:427	SEQ ID No:428
ILF1	179	interleukin enhancer binding factor 1	SEQ ID No:429		SEQ ID No:408
EST T80406	180	similar to SP:S36648 S36648 RB2/P130 PROTEIN	SEQ ID No:430		
EST T95640	181	similar to gb:M16336 T-CELL SURFACE ANTIGEN CD2	SEQ ID No:431		
EST R28523	182	similar to placental lactogen (CSH1)	SEQ ID No:432		
ESTs H21879 & H21880	183	Homo sapiens plasminogen activator (PLAT)	SEQ ID No:433	SEQ ID No:434	
ESTs H24628 & H24592	184	Homo sapiens aminoacylase 1 (ACY1).	SEQ ID No:435	SEQ ID No:436	
EST H28056	185	Homo sapiens E74-like factor 1 (ets domain transcription factor) (ELF1)	SEQ ID No:437		
ESTs H30141 & H27466	186	Homo sapiens selectin P	SEQ ID No:438	SEQ ID No:439	
ESTs H42957 & H42888	187	Human interleukin 3 receptor (hIL-3Ra)	SEQ ID No:440	SEQ ID No:441	
EST H57912	188	Human tumor protein p53 (Li-Fraumeni syndrome) (TP53)	SEQ ID No:442	SEQ ID No:443	
ERBB2	189	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2 (neuro/glioblastoma derived oncogene homolog) (ERBB2)	SEQ ID No:444		
ZNF144	190	zinc finger protein 144 (Mel-18) (ZNF144)	SEQ ID No:445		

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
MARK3	191	MAP/microtubule affinity-regulating kinase 3 (MARK3)	SEQ ID No:446	SEQ ID No:447	
EST N68536	192	EST N68536 MAX-interacting protein 1 (MXI1)	SEQ ID No:448		
EST R81126	193	EST R81126 lymphotoxin beta receptor (LTBR)		SEQ ID No:449	
POU2F2	194	(POU2F2)		SEQ ID No:450	
CASP1	195	caspase 4, apoptosis-related cysteine protease (CASP4) (ex CASP1)		SEQ ID No:451	
HRB	196	syndecan 1 (SDC1) (ex HRB)		SEQ ID No:452	
ITGB2	197	integrin, beta 2 (antigen CD18 (p95), lymphocyte function-associated antigen 1; macrophage antigen 1 (mac-1) beta subunit) (ITGB2)	SEQ ID No:453		
MGST1	198	protein phosphatase 1, catalytic subunit, alpha isoform (PPP1CA) (ex MGST1)		SEQ ID No:454	
PPP2CA	199	protein phosphatase 2 (formerly 2A), catalytic subunit,alpha isoform (PPP2CA)	SEQ ID No:455		
SUI1	200	S100 calcium-binding protein A11 (calgizzarin) (S100A11)		SEQ ID No:456	
GZMA	201	granzyme A (granzyme 1, cytotoxic T-lymphocyte-associated serine esterase 3) (GZMA)		SEQ ID No:457	
EDN1	202	endothelin 1 (EDN1)	SEQ ID No:458		
PTPN6	203	protein tyrosine phosphatase, non-receptor type 6 (PTPN6)	SEQ ID No:459		
TFAP4	204	transcription factor AP-4 (activating enhancer binding protein 4) (TFAP4)	SEQ ID No:460		
CCND2	205	cyclin D2 (CCND2)	SEQ ID No:461		
JUP	206	junction plakoglobin (JUP)	SEQ ID No:462		
GADD45A	207	growth arrest and DNA-damage-inducible, alpha (GADD45A)	SEQ ID No:463		
nm23	208	non-metastatic cells 1, protein (NM23A) expressed in (NME1)	SEQ ID No:464		
BBC1	209	ribosomal protein L13 (RPL13) (ex	SEQ ID		

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
		BBC1)	No:465		
VEGFB	210	vascular endothelial growth factor B (VEGFB)	SEQ ID No:466		
LAMR1	211	laminin receptor 1 (67kD, ribosomal protein SA) (LAMR1)	SEQ ID No:467		
CSH1	212	Chorionic somatomammotropin hormone 1 (placental lactogen) = LACTOGEN Precursor		SEQ ID No:468	

Tables 5A and 5B hereunder displays two subpopulations corresponding to the 5 top overexpressed and to the 5 top underexpressed polynucleotide sequences particularly interesting to distinguish healthy person from cancer patient.

TABLE 5A

overexpressed genes : top 5

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
GATA3	32	gata-binding protein 3	SEQ ID No:76	SEQ ID No:77	SEQ ID No:78
GZMB	73	granzyme b (granzyme 2, cytotoxic t-lymphocyte-associated serine esterase 1)	SEQ ID No:178		SEQ ID No:179
MYBL2	131	v-myb avian myeloblastosis viral oncogene homolog-like 2	SEQ ID No:308	SEQ ID No:309	SEQ ID No:310
MMP11	145	matrix metalloproteinase 11 (stromelysin 3)	SEQ ID No:345		SEQ ID No:346
EST T95640	181	similar to gb:M16336 T-CELL SURFACE ANTIGEN CD2	SEQ ID No:431		

TABLE 5B

underexpressed genes : top 5

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
PRL	38	prolactin	SEQ ID No:91	SEQ ID No:92	SEQ ID No:93
TEK	58	tek tyrosine kinase, endothelial (venous malformations, multiple cutaneous and mucosal)	SEQ ID No:138	SEQ ID No:139	SEQ ID No:140
PLA2G2A	61	phospholipase a2, group iia (platelets, synovial fluid)	SEQ ID No:147	SEQ ID No:148	SEQ ID No:149
DES	69	desmin	SEQ ID No:168	SEQ ID No:169	SEQ ID No:170
EST R28523	182	similar to placental lactogen (CSH1)	SEQ ID No:432		

Table 6 hereunder relate to sub populations of polynucleotide sequences interesting to detect hormone sensitive tumors allowing to distinguish between ER+ and ER- samples.

TABLE 6

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
SOX4	11	sry (sex determining region y)-box 4	SEQ ID No:22	SEQ ID No:23	SEQ ID No:24
IGF2	26	insulin-like growth factor 2 (somatomedin a)	SEQ ID No:59	SEQ ID No:60	SEQ ID No:61
GATA3	32	gata-binding protein 3	SEQ ID No:76	SEQ ID No:77	SEQ ID No:78
TOP2B	34	topoisomerase (dna) ii beta (180kd)		SEQ ID No:82	SEQ ID No:83
IL2RB	40	interleukin 2 receptor, beta	SEQ ID No:97	SEQ ID No:98	SEQ ID No:99
EGFR	57	epidermal growth factor receptor (avian erythroblastic leukemia viral (v-erb-b) oncogene homolog)	SEQ ID No:135	SEQ ID No:136	SEQ ID No:137
CRABP2	64	cellular retinoic acid-binding protein 2	SEQ ID No:156	SEQ ID No:157	SEQ ID No:158

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
S100B	107	s100 calcium-binding protein, beta (neural)		SEQ ID No:255	SEQ ID No:256
IL2RG	119	interleukin 2 receptor, gamma (severe combined immunodeficiency)	SEQ ID No:279	SEQ ID No:280	SEQ ID No:281
KIAA1075	136	kiaa1075 protein	SEQ ID No:322	SEQ ID No:323	
MST1	140	macrophage stimulating 1 (hepatocyte growth factor-like)	SEQ ID No:331	SEQ ID No:332	SEQ ID No:333
GSTP1	141	glutathione s-transferase pi	SEQ ID No:334	SEQ ID No:335	SEQ ID No:336
MMP11	145	matrix metalloproteinase 11 (stromelysin 3)	SEQ ID No:345		SEQ ID No:346
FLJ11307	148	hypothetical protein flj11307	SEQ ID No:352		SEQ ID No:353
MYB	149	v-myb avian myeloblastosis viral oncogene homolog		SEQ ID No:354	SEQ ID No:355
XBP1	162	x-box binding protein 1	SEQ ID No:385	SEQ ID No:386	SEQ ID No:387
SOX9	165	sry (sex determining region y)-box 9 (campomelic dysplasia, autosomal sex-reversal)	SEQ ID No:394		SEQ ID No:395
GZMA	169	granzyme a (granzyme 1, cytotoxic t-lymphocyte-associated serine esterase 3)	SEQ ID No:402		SEQ ID No:403
CD3G	174	cd3g antigen, gamma polypeptide (tit3 complex)	SEQ ID No:414	SEQ ID No:415	SEQ ID No:416
EST H57912	188	Human tumor protein p53 (Li-Fraumeni syndrome) (TP53)	SEQ ID No:442		

Tables 6A et 6B hereunder relate to two sub populations of polynucleotide sequences particularly interesting to detect hormone sensitive tumors allowing to distinguish between ER+ and ER- samples

Table 6A

overexpressed genes : top 5

ER + / ER -

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
GATA3	32	gata-binding protein 3	SEQ ID No:76	SEQ ID No:77	SEQ ID No:78
KIAA1075	136	kiaa1075 protein	SEQ ID No:322	SEQ ID No:323	
MMP11	145	matrix metalloproteinase 11 (stromelysin 3)	SEQ ID No:345		SEQ ID No:346
MYB	149	v-myb avian myeloblastosis viral oncogene homolog		SEQ ID No:354	SEQ ID No:355
GZMA	169	granzyme a (granzyme 1, cytotoxic t-lymphocyte-associated serine esterase 3)	SEQ ID No:402		SEQ ID No:403

Table 6B

underexpressed genes : top 5

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
SOX4	11	sry (sex determining region y)-box 4	SEQ ID No:22	SEQ ID No:23	SEQ ID No:24
IL2RB	40	interleukin 2 receptor, beta	SEQ ID No:97	SEQ ID No:98	SEQ ID No:99
EGFR	57	epidermal growth factor receptor (avian erythroblastic leukemia viral (v-erb-b) oncogene homolog)	SEQ ID No:135	SEQ ID No:136	SEQ ID No:137
IL2RG	119	interleukin 2 receptor, gamma (severe combined immunodeficiency)	SEQ ID No:279	SEQ ID No:280	SEQ ID No:281
CD3G	174	cd3g antigen, gamma polypeptide (tit3 complex)	SEQ ID No:414	SEQ ID No:415	SEQ ID No:416

Tables 7 hereunder relates to subpopulations of polynucleotide sequences interesting to distinguish tumors with lymphe node from tumors with no lymphe node.

TABLE 7

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
EST T89980	8	ests	SEQ ID No:16		
SOX4	11	sry (sex determining region y)-box 4	SEQ ID No:22	SEQ ID No:23	SEQ ID No:24
ENPP2	18	ectonucleotide pyrophosphatase/phosphodiesterase 2 (autotaxin)	SEQ ID No:39	SEQ ID No:40	SEQ ID No:41
MUC1	25	mucin 1, transmembrane		SEQ ID No:57	SEQ ID No:58
GATA3	32	gata-binding protein 3	SEQ ID No:76	SEQ ID No:77	SEQ ID No:78
TOP2B	34	topoisomerase (dna) ii beta (180kd)		SEQ ID No:82	SEQ ID No:83
IL2RB	40	interleukin 2 receptor, beta	SEQ ID No:97	SEQ ID No:98	SEQ ID No:99
ERBB2	49	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2 (neuro/glioblastoma derived oncogene homolog)		SEQ ID No:118	SEQ ID No:119
EGFR	57	epidermal growth factor receptor (avian erythroblastic leukemia viral (v-erb-b) oncogene homolog)	SEQ ID No:135	SEQ ID No:136	SEQ ID No:137
THBS1	91	thrombospondin 1	SEQ ID No:216		SEQ ID No:217
PPP2R2C	100	protein phosphatase 2 (formerly 2a), regulatory subunit b (pr 52), gamma isoform	SEQ ID No:238	SEQ ID No:239	
ATF3	105	activating transcription factor 3	SEQ ID No:250	SEQ ID No:251	SEQ ID No:252
KIAA1075	136	kiaa1075 protein	SEQ ID No:322	SEQ ID No:323	
CDH1	138	cadherin 1, type 1, e-cadherin (epithelial)	SEQ ID No:326	SEQ ID No:327	SEQ ID No:328
ZNF144	139	zinc finger protein 144 (mel-18)		SEQ ID No:329	SEQ ID No:330
GSTP1	141	glutathione s-transferase pi	SEQ ID No:334	SEQ ID No:335	SEQ ID No:336
CD44	158	cd44 antigen (homing function and indian blood group system)	SEQ ID No:374	SEQ ID No:375	SEQ ID No:376
GZMA	169	granzyme a (granzyme 1, cytotoxic t-lymphocyte-associated serine esterase 3)	SEQ ID No:402		SEQ ID No:403

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
EST T80406	180	similar to SP:S36648 S36648 RB2/P130 PROTEIN	SEQ ID No:430		
ESTs H30141 & H27466	186	Homo sapiens selectin P	SEQ ID No:438	SEQ ID No:439	

Tables 7A and 7B hereunder relate to two sub populations of polynucleotide sequences particularly interesting to distinguish tumors with lymphe node from tumors with no lymphe node.

TABLE 7A

Overexpressed genes : top 5

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
ENPP2	18	ectonucleotide pyrophosphatase/phosphodiestera se 2 (autotaxin)	SEQ ID No:39	SEQ ID No:40	SEQ ID No:41
GATA3	32	gata-binding protein 3	SEQ ID No:76	SEQ ID No:77	SEQ ID No:78
EGFR	57	epidermal growth factor receptor (avian erythroblastic leukemia viral (v-erb-b) oncogene homolog)	SEQ ID No:135	SEQ ID No:136	SEQ ID No:137
THBS1	91	thrombospondin 1	SEQ ID No:216		SEQ ID No:217
ATF3	105	activating transcription factor 3	SEQ ID No:250	SEQ ID No:251	SEQ ID No:252

TABLE 7B

Underexpressed genes : top 5

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
SOX4	11	sry (sex determining region y)-box 4	SEQ ID No:22	SEQ ID No:23	SEQ ID No:24
IL2RB	40	interleukin 2 receptor, beta	SEQ ID No:97	SEQ ID No:98	SEQ ID No:99

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
ERBB2	49	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2 (neuro/glioblastoma derived oncogene homolog)		SEQ ID No:118	SEQ ID No:119
PPP2R2C	100	protein phosphatase 2 (formerly 2a), regulatory subunit b (pr 52), gamma isoform	SEQ ID No:238	SEQ ID No:239	
GSTP1	141	glutathione s-transferase pi	SEQ ID No:334	SEQ ID No:335	SEQ ID No:336

Tables 8, 8A and 8B hereunder relates to sub populations of polynucleotide sequences particularly interesting to distinguish tumors sensitive to antracycline from tumors unsensitive to antracycline.

TABLE 8

A1 /A2

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
SOX4	11	sry (sex determining region y)-box 4	SEQ ID No:22	SEQ ID No:23	SEQ ID No:24
CSF1	22	colony stimulating factor 1 (macrophage)	SEQ ID No:48	SEQ ID No:49	SEQ ID No:50
VIL2	23	villin 2 (ezrin)	SEQ ID No:51	SEQ ID No:52	SEQ ID No:53
IGF2	26	insulin-like growth factor 2 (somatomedin a)	SEQ ID No:59	SEQ ID No:60	SEQ ID No:61
KIAA0427	28	kiaa0427 gene product	SEQ ID No:65	SEQ ID No:66	SEQ ID No:67
MYC	31	v-myc avian myelocytomatosis viral oncogene homolog	SEQ ID No:73	SEQ ID No:74	SEQ ID No:75
GATA3	32	gata-binding protein 3	SEQ ID No:76	SEQ ID No:77	SEQ ID No:78
TOP2B	34	topoisomerase (dna) ii beta (180kd)		SEQ ID No:82	SEQ ID No:83
ERBB2	49	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2 (neuro/glioblastoma derived oncogene homolog)		SEQ ID No:118	SEQ ID No:119
EGFR	57	epidermal growth factor receptor (avian erythroblastic leukemia viral (v-erb-b) oncogene homolog)	SEQ ID No:135	SEQ ID No:136	SEQ ID No:137

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
CRABP2	64	cellular retinoic acid-binding protein 2	SEQ ID No:156	SEQ ID No:157	SEQ ID No:158
GZMB	73	granzyme b (granzyme 2, cytotoxic t-lymphocyte-associated serine esterase 1)	SEQ ID No:178		SEQ ID No:179
IGKC	77	immunoglobulin kappa constant	SEQ ID No:186		
ANG	81	angiogenin, ribonuclease, rnase a family, 5		SEQ ID No:194	SEQ ID No:195
EFNA1	95	ephrin-a1		SEQ ID No:226	SEQ ID No:227
MYBL2	131	v-myb avian myeloblastosis viral oncogene homolog-like 2	SEQ ID No:308	SEQ ID No:309	SEQ ID No:310
CDH1	138	cadherin 1, type 1, e-cadherin (epithelial)	SEQ ID No:326	SEQ ID No:327	SEQ ID No:328
MST1	140	macrophage stimulating 1 (hepatocyte growth factor-like)	SEQ ID No:331	SEQ ID No:332	SEQ ID No:333
MYB	149	v-myb avian myeloblastosis viral oncogene homolog		SEQ ID No:354	SEQ ID No:355
XBP1	162	x-box binding protein 1	SEQ ID No:385	SEQ ID No:386	SEQ ID No:387
SRF	164	serum response factor (c-fos serum response element-binding transcription factor)	SEQ ID No:391	SEQ ID No:392	SEQ ID No:393
SOX9	165	sry (sex determining region y)-box 9 (campomelic dysplasia, autosomal sex-reversal)	SEQ ID No:394		SEQ ID No:395
ESTs H21879 & H21880	183	Homo sapiens plasminogen activator (PLAT)	SEQ ID No:433	SEQ ID No:434	

Tables 8A and 8B hereunder relate to two sub populations of polynucleotide sequences particularly interesting to distinguish tumors sensitive to antracycline from tumors unsensitive to antracycline.

TABLEAU 8A

overexpressed genes : top 5

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
GATA3	32	gata-binding protein 3	SEQ ID No:76	SEQ ID No:77	SEQ ID No:78
KIAA1075	136	kiaa1075 protein	SEQ ID No:322	SEQ ID No:323	
MMP11	145	matrix metalloproteinase 11 (stromelysin 3)	SEQ ID No:345		SEQ ID No:346
MYB	149	v-myb avian myeloblastosis viral oncogene homolog		SEQ ID No:354	SEQ ID No:355
GZMA	169	granzyme a (granzyme 1, cytotoxic t-lymphocyte-associated serine esterase 3)	SEQ ID No:402		SEQ ID No:403

TABLEAU 8B

underexpressed genes : top 5

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
SOX4	11	sry (sex determining region y)-box 4	SEQ ID No:22	SEQ ID No:23	SEQ ID No:24
IL2RB	40	interleukin 2 receptor, beta	SEQ ID No:97	SEQ ID No:98	SEQ ID No:99
EGFR	57	epidermal growth factor receptor (avian erythroblastic leukemia viral (v-erb-b) oncogene homolog)	SEQ ID No:135	SEQ ID No:136	SEQ ID No:137
IL2RG	119	interleukin 2 receptor, gamma (severe combined immunodeficiency)	SEQ ID No:279	SEQ ID No:280	SEQ ID No:281
CD3G	174	cd3g antigen, gamma polypeptide (tit3 complex)	SEQ ID No:414	SEQ ID No:415	SEQ ID No:416

Tables 9, 9A and 9B hereunder relates to sub populations of polynucleotide sequences particularly interesting in classifying good and poor prognosis primary breast tumors.

TABLE 9

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
CTSB	14	cathepsin b		SEQ ID No:30	SEQ ID No:31
VIL2	23	villin 2 (ezrin)	SEQ ID No:51	SEQ ID No:52	SEQ ID No:53
MUC1	25	mucin 1, transmembrane		SEQ ID No:57	SEQ ID No:58
EMR1	27	egf-like module containing, mucin-like, hormone receptor-like sequence 1	SEQ ID No:62	SEQ ID No:63	SEQ ID No:64
KIAA0427	28	kiaa0427 gene product	SEQ ID No:65	SEQ ID No:66	SEQ ID No:67
GATA3	32	gata-binding protein 3	SEQ ID No:76	SEQ ID No:77	SEQ ID No:78
PRLR	39	prolactin receptor	SEQ ID No:94	SEQ ID No:95	SEQ ID No:96
GATA3	41	gata-binding protein 3	SEQ ID No:100	SEQ ID No:101	SEQ ID No:78
TC21	44	oncogene tc21	SEQ ID No:106	SEQ ID No:107	SEQ ID No:108
BCL2	48	b-cell cll/lymphoma 2	SEQ ID No:115	SEQ ID No:116	SEQ ID No:117
GATA3	51	gata-binding protein 3	SEQ ID No:122		SEQ ID No:78
CRABP2	64	cellular retinoic acid-binding protein 2	SEQ ID No:156	SEQ ID No:157	SEQ ID No:158
ANG	81	angiogenin, ribonuclease, rnase a family, 5		SEQ ID No:194	SEQ ID No:195
EGF	83	epidermal growth factor (beta-urogastrone)	SEQ ID No:199		SEQ ID No:200
THBS1	91	thrombospondin 1	SEQ ID No:216		SEQ ID No:217
EDNRA	96	endothelin receptor type a	SEQ ID No:228		SEQ ID No:229
SMARCA2	99	swi/snf related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2	SEQ ID No:235	SEQ ID No:236	SEQ ID No:237
ABCB1	108	atp-binding cassette, sub-family b (mdr/tap), member 1	SEQ ID No:257		SEQ ID No:258
EGF	110	epidermal growth factor (beta-urogastrone)	SEQ ID No:262		SEQ ID No:200
BIRC4	116	baculoviral iap repeat-containing 4	SEQ ID No:273		SEQ ID No:274
DAP3	117	death associated protein 3	SEQ ID		SEQ ID

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
			No:275		No:276
GNRH1	118	gonadotropin-releasing hormone 1 (leutinizing-releasing hormone)		SEQ ID No:277	SEQ ID No:278
DAP3	120	death associated protein 3	SEQ ID No:282	SEQ ID No:283	SEQ ID No:276
EST R97218	126	ests, highly similar to tvhume hepatocyte growth factor receptor precursor [h.sapiens]	SEQ ID No:296	SEQ ID No:297	
BCL2	142	b-cell cll/lymphoma 2	SEQ ID No:337	SEQ ID No:338	SEQ ID No:117
BS69	144	adenovirus 5 e1a binding protein	SEQ ID No:342	SEQ ID No:343	SEQ ID No:344
MYB	149	v-myb avian myeloblastosis viral oncogene homolog		SEQ ID No:354	SEQ ID No:355
CTSB	152	cathepsin b	SEQ ID No:361		SEQ ID No:31
MLANA	153	melan-a	SEQ ID No:362	SEQ ID No:363	SEQ ID No:364
APR-1	154	apr-1 protein	SEQ ID No:365	SEQ ID No:366	SEQ ID No:367
TC21	157	oncogene tc21	SEQ ID No:372	SEQ ID No:373	SEQ ID No:108
CDKN3	159	cyclin-dependent kinase inhibitor 3 (cdk2-associated dual specificity phosphatase)	SEQ ID No:377	SEQ ID No:378	SEQ ID No:379
XBP1	162	x-box binding protein 1	SEQ ID No:385	SEQ ID No:386	SEQ ID No:387
CDH15	166	cadherin 15, m-cadherin (myotubule)	SEQ ID No:396	SEQ ID No:397	SEQ ID No:398
BCL2	167	b-cell cll/lymphoma 2	SEQ ID No:399	SEQ ID No:400	SEQ ID No:117
EST W73386	168	ests	SEQ ID No:401		
ILF1	171	interleukin enhancer binding factor 1	SEQ ID No:406	SEQ ID No:407	SEQ ID No:408
ARHGDIA	172	rho gdp dissociation inhibitor (gdi) alpha	SEQ ID No:409	SEQ ID No:410	SEQ ID No:411
C4A	173	complement component 4a	SEQ ID No:412		SEQ ID No:413
ESR1	176	estrogen receptor 1	SEQ ID No:420	SEQ ID No:421	SEQ ID No:422
PBX1	177	pre-b-cell leukemia transcription factor 1	SEQ ID No:423	SEQ ID No:424	SEQ ID No:425
GLI3	178	gli-kruppel family member gli3 (greig	SEQ ID No:426	SEQ ID No:427	SEQ ID No:428

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
		cephalopolysyndactyly syndrome)			
ILF1	179	interleukin enhancer binding factor 1	SEQ ID No:429		SEQ ID No:408
ESTs H24628 & H24592	184	Homo sapiens aminoacylase 1 (ACY1).	SEQ ID No:435	SEQ ID No:436	
EST H28056	185	Homo sapiens E74-like factor 1 (ets domain transcription factor) (ELF1)	SEQ ID No:437		

TABLE 9A

Gene symbol	SET N°	Name	Seq3'	Seq5'	Ref
VIL2	23	villin 2 (ezrin)	SEQ ID No:51	SEQ ID No:52	SEQ ID No:53
MUC1	25	mucin 1, transmembrane		SEQ ID No:57	SEQ ID No:58
GATA3	32	gata-binding protein 3	SEQ ID No:76	SEQ ID No:77	SEQ ID No:78
GATA3	41	gata-binding protein 3	SEQ ID No:100	SEQ ID No:101	SEQ ID No:78
BCL2	48	b-cell cll/lymphoma 2	SEQ ID No:115	SEQ ID No:116	SEQ ID No:117
GATA3	51	gata-binding protein 3	SEQ ID No:122		SEQ ID No:78
CRABP2	64	cellular retinoic acid-binding protein 2	SEQ ID No:156	SEQ ID No:157	SEQ ID No:158
ANG	81	angiogenin, ribonuclease, rnase a family, 5		SEQ ID No:194	SEQ ID No:195
EGF	83	epidermal growth factor (beta-urogastrone)	SEQ ID No:199		SEQ ID No:200
THBS1	91	thrombospondin 1	SEQ ID No:216		SEQ ID No:217
SMARCA2	99	swi/snf related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2	SEQ ID No:235	SEQ ID No:236	SEQ ID No:237
EGF	110	epidermal growth factor (beta-urogastrone)	SEQ ID No:262		SEQ ID No:200
BIRC4	116	baculoviral iap repeat-containing 4	SEQ ID No:273		SEQ ID No:274
BCL2	142	b-cell cll/lymphoma 2	SEQ ID No:337	SEQ ID No:338	SEQ ID No:117

Gene symbol	SET N°	Name	Seq3'	Seq5'	Ref
BS69	144	adenovirus 5 e1a binding protein	SEQ ID No:342	SEQ ID No:343	SEQ ID No:344
MYB	149	v-myb avian myeloblastosis viral oncogene homolog		SEQ ID No:354	SEQ ID No:355
XBP1	162	x-box binding protein 1	SEQ ID No:385	SEQ ID No:386	SEQ ID No:387
BCL2	167	b-cell cll/lymphoma 2	SEQ ID No:399	SEQ ID No:400	SEQ ID No:117
ILF1	171	interleukin enhancer binding factor 1	SEQ ID No:406	SEQ ID No:407	SEQ ID No:408
ARHGDIA	172	rho gdp dissociation inhibitor (gdi) alpha	SEQ ID No:409	SEQ ID No:410	SEQ ID No:411
C4A	173	complement component 4a	SEQ ID No:412		SEQ ID No:413
ESR1	176	estrogen receptor 1	SEQ ID No:420	SEQ ID No:421	SEQ ID No:422
PBX1	177	pre-b-cell leukemia transcription factor 1	SEQ ID No:423	SEQ ID No:424	SEQ ID No:425
GLI3	178	gli-kruppel family member gli3 (greig cephalopolysyndactyly syndrome)	SEQ ID No:426	SEQ ID No:427	SEQ ID No:428
ILF1	179	interleukin enhancer binding factor 1	SEQ ID No:429		SEQ ID No:408
ESTs H24628 & H24592	184	Homo sapiens aminoacylase 1 (ACY1).	SEQ ID No:435	SEQ ID No:436	
EST H28056	185	Homo sapiens E74-like factor 1 (ets domain transcription factor) (ELF1)	SEQ ID No:437		

TABLE 9B

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
CTSB	14	cathepsin b		SEQ ID No:30	SEQ ID No:31
EMR1	27	egf-like module containing, mucin-like, hormone receptor-like sequence 1	SEQ ID No:62	SEQ ID No:63	SEQ ID No:64
KIAA0427	28	kiaa0427 gene product	SEQ ID No:65	SEQ ID No:66	SEQ ID No:67
PRLR	39	prolactin receptor	SEQ ID No:94	SEQ ID No:95	SEQ ID No:96
TC21	44	oncogene tc21	SEQ ID No:106	SEQ ID No:107	SEQ ID No:108

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
EDNRA	96	endothelin receptor type a	SEQ ID No:228		SEQ ID No:229
ABCB1	108	atp-binding cassette, sub-family b (mdr/tap), member 1	SEQ ID No:257		SEQ ID No:258
DAP3	117	death associated protein 3	SEQ ID No:275		SEQ ID No:276
GNRH1	118	gonadotropin-releasing hormone 1 (leutinizing-releasing hormone)		SEQ ID No:277	SEQ ID No:278
DAP3	120	death associated protein 3	SEQ ID No:282	SEQ ID No:283	SEQ ID No:276
EST R97218	126	ests, highly similar to tvhume hepatocyte growth factor receptor precursor [h.sapiens]	SEQ ID No:296	SEQ ID No:297	
CTSB	152	cathepsin b	SEQ ID No:361		SEQ ID No:31
MLANA	153	melan-a	SEQ ID No:362	SEQ ID No:363	SEQ ID No:364
APR-1	154	apr-1 protein	SEQ ID No:365	SEQ ID No:366	SEQ ID No:367
TC21	157	oncogene tc21	SEQ ID No:372	SEQ ID No:373	SEQ ID No:108
CDKN3	159	cyclin-dependent kinase inhibitor 3 (cdk2-associated dual specificity phosphatase)	SEQ ID No:377	SEQ ID No:378	SEQ ID No:379
CDH15	166	cadherin 15, m-cadherin (myotubule)	SEQ ID No:396	SEQ ID No:397	SEQ ID No:398
EST W73386	168	ests	SEQ ID No:401		

Overexpression of genes detected by using at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences indicated in table 9A combined with underexpression of genes detected with at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequence indicated on table 9B present a Good outcome.

So, a preferred DNA array according to the invention comprises at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences indicated in table 9A and at least

one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequence indicated on table 9B.

Such DNA arrays are particularly useful to distinguish patients having a high risk (Bad Outcome) from those having a good pronostic (Good Outcome).

TABLE 10

CORRELATION BETWEEN SEQ ID NO AS FILED WITH US PROVISIONAL APPLICATION N° 60/254,090

and SEQ ID NO FILED WITH PCT APPLICATION

Symbole gène	N°	Nom	Image	Seq3' US PROV LISTING	Seq5' US PROV LISTING		Seq3' PCT Listing	Seq5' PCT Listing	(mRNA) PCT Listing
GATA3	1	GATA-binding protein 3 (GATA3)	129757	SEQ ID No : 1			SEQ ID No:76	SEQ ID No:77	SEQ ID No:78
MYB	2	v-myb avian myeloblastosis viral oncogene homolog (MYB)	248613		SEQ ID No : 2		0	SEQ ID No:354	SEQ ID No:355
KIAA1075	3	KIAA1075 protein	211999	SEQ ID No : 3	SEQ ID No : 4		SEQ ID No:322	SEQ ID No:323	0
STMY3	4	matrix metalloproteinase 11 (stromelysin 3) (MMP11) (ex STMY3)	235947	SEQ ID No : 5			SEQ ID No:345	0	SEQ ID No:346
HGFL	5	macrophage-stimulating protein (MST1) (ex HGFL)	229839	SEQ ID No : 6	SEQ ID No : 7		SEQ ID No:331	SEQ ID No:332	SEQ ID No:333
CRABP	6	cellular retinoic acid-binding protein 2 (CRABP2)	153275	SEQ ID No : 8	SEQ ID No : 9		SEQ ID No:156	SEQ ID No:157	SEQ ID No:158
XBP1	7	X-box binding protein 1 (XBP1)	301950	SEQ ID No : 10	SEQ ID No : 11		SEQ ID No:385	SEQ ID No:386	SEQ ID No:387
TP53	8	tumor protein p53 (Li-Fraumeni syndrome) (TP53)	205314		SEQ ID No : 12		SEQ ID No:442	0	0
IGF2	9	insulin-like growth factor 2 (somatomedin A) (IGF2),	126233	SEQ ID No : 13	SEQ ID No : 14		SEQ ID No:59	SEQ ID No:60	SEQ ID No:61
CD3G	10	CD3G antigen, gamma polypeptide (TiT3 complex) (CD3G)	66322	SEQ ID No : 15	SEQ ID No : 16		SEQ ID No:414	SEQ ID No:415	SEQ ID No:416
IL2RG	11	interleukin 2 receptor, gamma (severe combined immunodeficiency) (IL2RG)	195022	SEQ ID No : 17	SEQ ID No : 18		SEQ ID No:279	SEQ ID No:280	SEQ ID No:281
SOX4	12	SRY (sex determining region Y)-box 4 (SOX4)	111461	SEQ ID No : 19	SEQ ID No : 20		SEQ ID No:22	SEQ ID No:23	SEQ ID No:24
EGFR	13	epidermal growth factor receptor (avian erythroblastic	151475	SEQ ID No : 21	SEQ ID No : 22		SEQ ID No:135	SEQ ID No:136	SEQ ID No:137
TOP2B	14	topIIb mRNA for topoisomerase IIb.	130788		SEQ ID No : 23		0	SEQ ID No:82	SEQ ID No:83
S100B	15	S100 calcium-binding protein, beta (neural) (S100B)	183641		SEQ ID No : 24		0	SEQ ID No:255	SEQ ID No:256
EST N53133	16	EST N53133	246620	SEQ ID No : 25			SEQ ID No:352	0	SEQ ID No:353
GSTP1	17	glutathione S-transferase pi (GSTP1)	231424	SEQ ID No : 26	SEQ ID No : 27		SEQ ID No:334	SEQ ID No:335	SEQ ID No:336
THBS1	18	thrombospondin 1 (THBS1)	160963	SEQ ID No : 28			SEQ ID No:216	0	SEQ ID No:217

Symbole gène	N°	Nom	Image	Seq3' US PROV LISTING	Seq5' US PROV LISTING	Seq3' PCT Listing	Seq5' PCT Listing	(mRNA) PCT Listing
PDNP2	19	ectonucleotide pyrophosphatase/phosphodiesterase 2(autotaxin) (ENPP2) (ex PDNP2)	120916	SEQ ID No : 29	SEQ ID No : 30	SEQ ID No:39	SEQ ID No:40	SEQ ID No:41
ATF3	20	activating transcription factor 3 (ATF3)	183030	SEQ ID No : 31	SEQ ID No : 32	SEQ ID No:250	SEQ ID No:251	SEQ ID No:252
NF1A	21	(ex NF1A)	110480	SEQ ID No : 33		SEQ ID No:16	0	0
SELP	22	selectin P (granule membrane protein 140kD, antigen CD62) (SELP)	182264		SEQ ID No : 34	SEQ ID No:438	SEQ ID No:439	0
CDH1	23	cadherin 1, E-cadherin (epithelial) (CDH1)	214008	SEQ ID No : 35	SEQ ID No : 36	SEQ ID No:326	SEQ ID No:327	SEQ ID No:328
ERBB2	24	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2 (neuro/glioblastoma derived oncogene homolog) (ERBB2)	147016	SEQ ID No : 37		0	SEQ ID No:118	SEQ ID No:119
PP2A BR gamma	25	(PP2A BR gamma)	179197	SEQ ID No : 38	SEQ ID No : 39	SEQ ID No:238	SEQ ID No:239	0
ZNF144	26	zinc finger protein 144 (Mel-18) (ZNF144)	220451	SEQ ID No : 40	SEQ ID No : 41	0	SEQ ID No:329	SEQ ID No:330
MUC1	27	mucin 1, transmembrane (MUC1)	125413		SEQ ID No : 42	0	SEQ ID No:57	SEQ ID No:58
CD44	28	CD44E (epithelial form)	290007	SEQ ID No : 43	SEQ ID No : 44	SEQ ID No:374	SEQ ID No:375	SEQ ID No:376
PLA2G2A	29	phospholipase A2, group IIA (platelets, synovial fluid) (PLA2G2A), nuclear gene encoding mitochondrial protein	152802	SEQ ID No : 45	SEQ ID No : 46	SEQ ID No:147	SEQ ID No:148	SEQ ID No:149
ACVRL1	30	activin A receptor type II-like 1 (ACVRL1)	153350	SEQ ID No : 47	SEQ ID No : 48	SEQ ID No:159	SEQ ID No:160	SEQ ID No:161
AXL	31	AXL receptor tyrosine kinase (AXL)	112500	SEQ ID No : 49	SEQ ID No : 50	SEQ ID No:27	SEQ ID No:28	SEQ ID No:29
PKU-ALPHA	32	KU-alpha, partial cds (new gene symbol Tlk2)	109569		SEQ ID No : 51	0	SEQ ID No:5	SEQ ID No:6
ABCC5	33	ATP-binding cassette, sub-family C (CFTR/MRP), member 5 (ABCC5)	212366		SEQ ID No : 52	0	SEQ ID No:324	SEQ ID No:325
EDNRB	34	endothelin receptor type B (EDNRB), transcript variant1	154244		SEQ ID No : 53	0	SEQ ID No:176	SEQ ID No:177
DTR	35	diphtheria toxin receptor (heparin-binding epidermal	187547		SEQ ID No : 54	0	SEQ ID No:265	SEQ ID No:266
IGF1R	36	insulin-like growth factor 1 receptor (IGF1R)	150361		SEQ ID No : 55	0	SEQ ID No:129	SEQ ID No:130
KIAA0427	37	KIAA0427	127507	SEQ ID No : 56	SEQ ID No : 57	SEQ ID No:65	SEQ ID No:66	SEQ ID No:67
CD69	38	CD69 antigen (p60, early T-cell activation antigen)	276727		SEQ ID No : 58	0	SEQ ID No:370	SEQ ID No:371
FGFR4	39	fibroblast growth factor receptor 4 (FGFR4)	116781	SEQ ID No : 59	SEQ ID No : 60	SEQ ID No:36	SEQ ID No:37	SEQ ID No:38
EST T85683	40	EST T85683 cathepsin B (CTSB)	112622		SEQ ID No : 61	0	SEQ ID No:30	SEQ ID No:31
EST R00569	41	EST R00569 IL2-inducible T-cell kinase (ITK)	123871		SEQ ID No : 62	0	SEQ ID No:44	SEQ ID No:45

Symbole gène	N°	Nom	Image	Seq3' US PROV LISTING	Seq5' US PROV LISTING	Seq3' PCT Listing	Seq5' PCT Listing	(mRNA) PCT Listing
TGFBR3	42	transforming growth factor, beta receptor III (TGFBR3)	208118	SEQ ID No : 63	SEQ ID No : 64	SEQ ID No:311	SEQ ID No:312	SEQ ID No:313
INSR	43	insulin receptor (INSR)	151149		SEQ ID No : 65	0	SEQ ID No:131	SEQ ID No:132
MARK3	44	MAP/microtubule affinity-regulating kinase 3 (MARK3)	110599	SEQ ID No : 66	SEQ ID No : 67	#N/A	#N/A	#N/A
TIMP2	45	tissue inhibitor of metalloproteinase 2 (TIMP2)	131504		SEQ ID No : 68	0	SEQ ID No:86	SEQ ID No:87
EST R85557	46	EST R85557 thrombospondin 3 (THBS3)	180219	SEQ ID No : 69		SEQ ID No:240	0	SEQ ID No:241
GNRH1	47	gonadotropin-releasing hormone 1 (GNRH1)	192688		SEQ ID No : 70	0	SEQ ID No:277	SEQ ID No:278
FGFR2	48	fibroblast growth factor receptor 2 (FGFR2)	110387	SEQ ID No : 71	SEQ ID No : 72	SEQ ID No:13	SEQ ID No:14	SEQ ID No:15
NFKB2	49	NFKB2	114879	SEQ ID No : 73		SEQ ID No:35	0	0
VIL2	50	villin 2 (ezrin) (VIL2)	124701	SEQ ID No : 74	SEQ ID No : 75	SEQ ID No:51	SEQ ID No:52	SEQ ID No:53
ENG	51	endoglin (ENG)	156979	SEQ ID No : 76	SEQ ID No : 77	SEQ ID No:196	SEQ ID No:197	SEQ ID No:198
EPHA2	52	EphA2 (EPHA2)	162004	SEQ ID No : 78		SEQ ID No:221	0	SEQ ID No:222
CREM	53	cAMP responsive element modulator (CREM)	258584	SEQ ID No : 79	SEQ ID No : 80	SEQ ID No:358	SEQ ID No:359	SEQ ID No:360
ETV5-a	54	ets variant gene 5 (ETV5)	270549	SEQ ID No : 81	SEQ ID No : 82	SEQ ID No:368	SEQ ID No:369	SEQ ID No:300
EST N68536	55	EST N68536 MAX-interacting protein 1 (MXI1)	298242	SEQ ID No : 83	SEQ ID No : 84	0	SEQ ID No:380	SEQ ID No:381
EST R81126	56	EST R81126 lymphotoxin beta receptor (LTBR)	146635	SEQ ID No : 85	SEQ ID No : 86	SEQ ID No:114	0	0
POU2F2	57	(POU2F2)	188393	SEQ ID No : 87	SEQ ID No : 88	SEQ ID No:271	0	SEQ ID No:272
FLI1	58	Friend leukemia virus integration 1 (FLI1)	198144	SEQ ID No : 89	SEQ ID No : 90	SEQ ID No:293	SEQ ID No:294	SEQ ID No:295
TIE	59	tyrosine kinase with immunoglobulin and epidermal growth factor homology domains (TIE)	144081		SEQ ID No : 91	0	SEQ ID No:109	SEQ ID No:110
PRLR	60	prolactin receptor (PRLR)	138788	SEQ ID No : 92	SEQ ID No : 93	SEQ ID No:94	SEQ ID No:95	SEQ ID No:96
PPP3CA	61	protein phosphatase 3 (formerly 2B), catalytic subunit, gamma isoform (calcineurin A gamma) (PPP3CC) (ex PPP3CA)	110481	SEQ ID No : 94	SEQ ID No : 95	SEQ ID No:17	SEQ ID No:18	SEQ ID No:19
PTPN2	62	protein tyrosine phosphatase, non-receptor type 2 (PTPN2)	161451	SEQ ID No : 96	SEQ ID No : 97	SEQ ID No:218	SEQ ID No:219	SEQ ID No:220
PGF	63	placental growth factor, vascular endothelial growth factor-related protein (PGF)	139326		SEQ ID No : 98	0	SEQ ID No:102	SEQ ID No:103
TNFAIP3	64	tumor necrosis factor, alpha-induced	309943	SEQ ID No : 99		SEQ ID No:388	SEQ ID No:389	SEQ ID No:390

Symbole gène	N°	Nom	Image	Seq3' US PROV LISTING	Seq5' US PROV LISTING		Seq3' PCT Listing	Seq5' PCT Listing	(mRNA) PCT Listing
		protein 3 (TNFAIP3)							
PHB	65	PHB (prohibitin)	236008	SEQ ID No : 100			SEQ ID No:347	SEQ ID No:348	SEQ ID No:349
RIL	66	LIM domain protein (RIL)	153446		SEQ ID No : 101		0	SEQ ID No:162	SEQ ID No:163
MYBL2	67	v-myb avian myeloblastosis viral oncogene homolog-like 2 (MYBL2)	207378	SEQ ID No : 102	SEQ ID No : 103		SEQ ID No:308	SEQ ID No:309	SEQ ID No:310
RELB	68	v-rel avian reticuloendotheliosis viral oncogene homolog B (nuclear factor of kappa light polypeptide gene enhancer in B-cells 3) (RELB)	66969	SEQ ID No : 104	SEQ ID No : 105		SEQ ID No:417	SEQ ID No:418	SEQ ID No:419
EST R97218	69	Est R97218	200394	SEQ ID No : 106			SEQ ID No:296	SEQ ID No:297	0
GZMH	70	granzyme B (granzyme 2, cytotoxic T-lymphocyte-associated serine esterase 1) (GZMB) (ex GZMH)	154343	SEQ ID No : 107			SEQ ID No:178	0	SEQ ID No:179
MYC	71	c-myc proto-oncogene	129438	SEQ ID No : 108	SEQ ID No : 109		SEQ ID No:73	SEQ ID No:74	SEQ ID No:75
CASP1	72	caspase 4, apoptosis-related cysteine protease (CASP4) (ex CASP1)	131502		SEQ ID No : 110		SEQ ID No:84	0	SEQ ID No:85
SYK	73	spleen tyrosine kinase (SYK)	128142	SEQ ID No : 111	SEQ ID No : 112		SEQ ID No:68	SEQ ID No:69	SEQ ID No:70
EST H27202	74	EST H27202 transcription factor E1AF gene	158347	SEQ ID No : 113	SEQ ID No : 114		SEQ ID No:204	SEQ ID No:205	0
HRB	75	syndecan 1 (SDC1) (ex HRB)	108490	SEQ ID No : 115	SEQ ID No : 116		SEQ ID No:1	0	SEQ ID No:2
SHC1	76	p66shc (SHC)	153548		SEQ ID No : 117		0	SEQ ID No:164	SEQ ID No:165
CSF1	77	colony stimulating factor 1 (CSF1)	124554	SEQ ID No : 118	SEQ ID No : 119		SEQ ID No:48	SEQ ID No:49	SEQ ID No:50
UBE3A	78	ubiquitin protein ligase E3A (UBE3A)	141924		SEQ ID No : 120		0	SEQ ID No:104	SEQ ID No:105
FKHR	79	forkhead box O1A (rhabdomyosarcoma) (FOXO1A) (ex FKHR)	151247		SEQ ID No : 121		0	SEQ ID No:133	SEQ ID No:134
CSF1R	80	colony stimulating factor 1 receptor (CSF1R)	196282	SEQ ID No : 122			SEQ ID No:291	0	SEQ ID No:292
IFI75	81	interferon-induced protein 75 (IFI75)	205612	SEQ ID No : 123	SEQ ID No : 124		SEQ ID No:305	SEQ ID No:306	SEQ ID No:307
GATA1	82	GATA-binding protein 1 (globin transcription factor 1) (GATA1)	109093		SEQ ID No : 125		0	SEQ ID No:3	SEQ ID No:4
STAT1	83	signal transducer and activator of transcription 1 (STAT1)	110101		SEQ ID No : 126		0	SEQ ID No:11	SEQ ID No:12
CREBBP	84	CREB binding protein (Rubinstein-Taybi syndrome) (CREBBP)	109677	SEQ ID No : 127	SEQ ID No : 128		SEQ ID No:7	SEQ ID No:8	0
IL7R	85	interleukin 7 receptor (IL7R)	129059		SEQ ID No : 129		0	SEQ ID No:71	SEQ ID No:72
ANXA7	86	annexin A7 (ANXA7)	160580		SEQ ID No : 130		0	SEQ ID No:214	SEQ ID No:215

Symbole gène	N°	Nom	Image	Seq3' US PROV LISTING	Seq5' US PROV LISTING	Seq3' PCT Listing	Seq5' PCT Listing	(mRNA) PCT Listing
TNXA	87	tenascin XA (TNXA)	124340		SEQ ID No : 131	0	SEQ ID No:46	SEQ ID No:47
CNBP1	88	zinc finger protein 9 (a cellular retroviral nucleic acid binding protein) (ZNF9) (ex CNBP1)	251963	SEQ ID No : 132		SEQ ID No:356	0	SEQ ID No:357
CDK4-a	89	cyclin-dependent kinase 4 (CDK4)	204586	SEQ ID No : 133	SEQ ID No : 134	SEQ ID No:301	SEQ ID No:302	SEQ ID No:288
CSNK2B	90	gene for casein kinase II subunit beta (EC 2.7.1.37).	153879		SEQ ID No : 135	0	SEQ ID No:171	SEQ ID No:172
EFNA1	91	ephrin-A1 (EFNA1)	162997		SEQ ID No : 136	0	SEQ ID No:226	SEQ ID No:227
SELE	92	selectin E (endothelial adhesion molecule 1) (SELE)	186132	SEQ ID No : 137	SEQ ID No : 138	SEQ ID No:259	SEQ ID No:260	SEQ ID No:261
APC	93	adenomatosis polyposis coli (APC)	125294	SEQ ID No : 139	SEQ ID No : 140	SEQ ID No:54	SEQ ID No:55	SEQ ID No:56
FAK	94	PTK2 protein tyrosine kinase 2 (PTK2) (ex FAK)	195731		SEQ ID No : 141	0	SEQ ID No:284	SEQ ID No:285
FOS-a	95	v-fos FBJ murine osteosarcoma viral oncogene homolog (FOS)	208717		SEQ ID No : 142	0	SEQ ID No:317	SEQ ID No:318
FGFR1	96	fibroblast growth factor receptor (FGFr)	154472	SEQ ID No : 143	SEQ ID No : 144	SEQ ID No:180	SEQ ID No:181	SEQ ID No:182
MC1R	97	melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor) (MC1R)	155691		SEQ ID No : 145	0	SEQ ID No:187	SEQ ID No:188
PCNA	98	proliferating cell nuclear antigen (PCNA)	232941	SEQ ID No : 146	SEQ ID No : 147	SEQ ID No:339	SEQ ID No:340	SEQ ID No:341
DDT	99	D-dopachrome tautomerase (DDT)	132109	SEQ ID No : 148	SEQ ID No : 149	SEQ ID No:88	SEQ ID No:89	SEQ ID No:90
GRB2	100	growth factor receptor-bound protein 2 (GRB2)	172152	SEQ ID No : 150	SEQ ID No : 151	SEQ ID No:230	SEQ ID No:231	SEQ ID No:232
AMFR	101	autocrine motility factor receptor (AMFR)	146280	SEQ ID No : 152	SEQ ID No : 153	SEQ ID No:111	SEQ ID No:112	SEQ ID No:113
ITGB2	102	integrin, beta 2 (antigen CD18 (p95), lymphocyte function-associated antigen 1; macrophage antigen 1 (mac-1) beta subunit) (ITGB2)	187822	SEQ ID No : 154		0	SEQ ID No:267	SEQ ID No:268
JUND	103	jun D proto-oncogene (JUND)	175421	SEQ ID No : 155		SEQ ID No:233	0	SEQ ID No:234
NF45	104	interleukin enhancer binding factor 2 (ILF2) (ex NF45)	243907		SEQ ID No : 156	0	SEQ ID No:350	SEQ ID No:351
PPP4C	105	protein phosphatase 4 (formerly X) (PPP4C)	114097	SEQ ID No : 157	SEQ ID No : 158	SEQ ID No:32	SEQ ID No:33	SEQ ID No:34
EMS1	106	ATX1 (antioxidant protein 1, yeast) homolog 1 (ATOX1) (ex EMS1)	149172	SEQ ID No : 159		SEQ ID No:123	SEQ ID No:124	SEQ ID No:125
BCL2	107	B-cell CLL/lymphoma 2 (BCL2), nuclear gene encoding mitochondrial protein, transcript variant alpha	147002	SEQ ID No : 160	SEQ ID No : 161	SEQ ID No:115	SEQ ID No:116	SEQ ID No:117
MGST1	108	protein phosphatase 1, catalytic subunit,	182610	SEQ ID No : 162	SEQ ID No : 163	SEQ ID No:248	0	SEQ ID No:249

Symbole gène	N°	Nom	Image	Seq3' US PROV LISTING	Seq5' US PROV LISTING	Seq3' PCT Listing	Seq5' PCT Listing	(mRNA) PCT Listing
		alpha isoform (PPP1CA) (ex MGST1)						
PDGFRB	109	platelet-derived growth factor receptor, beta polypeptide (PDGFRB)	158976		SEQ ID No : 164	0	SEQ ID No:208	SEQ ID No:209
ANXA11	110	annexin A11 (ANXA11)	158892		SEQ ID No : 165	0	SEQ ID No:206	SEQ ID No:207
GPX1	111	histocompatibility class II antigen gamma chain (CD74) (ex GPX1 Glutation S transférase)	159809		SEQ ID No : 166	0	SEQ ID No:212	SEQ ID No:213
CFR-1	112	Golgi apparatus protein 1 (GLG1) (ex CFR-1)	153974	SEQ ID No : 167	SEQ ID No : 168	SEQ ID No:173	SEQ ID No:174	SEQ ID No:175
BTF3L3	113	basic transcription factor 3 (BTF3)	195889	SEQ ID No : 169		SEQ ID No:289	0	SEQ ID No:290
EST R55460	114	EST R55460	154997		SEQ ID No : 170	0	SEQ ID No:185	0
AKT2	115	v-akt murine thymoma viral oncogene homolog 2 (AKT2)	183552	SEQ ID No : 171		SEQ ID No:253	0	SEQ ID No:254
CDKN1A	116	cyclin-dependent kinase inhibitor (CDKN1A)	152524	SEQ ID No : 172	SEQ ID No : 173	SEQ ID No:144	SEQ ID No:145	SEQ ID No:146
PPP2CA	117	protein phosphatase 2 (formerly 2A), catalytic subunit,alpha isoform (PPP2CA)	154685	SEQ ID No : 174	SEQ ID No : 175	0	SEQ ID No:183	SEQ ID No:184
MDM2	118	mouse double minute 2, human homolog of; p53-binding protein (MDM2), transcript variant MDM2	148052	SEQ ID No : 176		0	SEQ ID No:120	SEQ ID No:121
TNFRSF6	119	tumor necrosis factor receptor superfamily, member 6 (TNFRSF6)	151767	SEQ ID No : 177	SEQ ID No : 178	SEQ ID No:141	SEQ ID No:142	SEQ ID No:143
CNTFR	120	ciliary neurotrophic factor receptor (CNTFR)	156431		SEQ ID No : 179	0	SEQ ID No:192	SEQ ID No:193
JUNB	121	jun B proto-oncogene (JUNB)	153213	SEQ ID No : 180	SEQ ID No : 181	SEQ ID No:153	SEQ ID No:154	SEQ ID No:155
CCND1	122	cyclin D1 (PRAD1: parathyroid adenomatosis 1) (CCND1)	110022	SEQ ID No : 182		SEQ ID No:9	0	SEQ ID No:10
TDPX1	123	peroxiredoxin 2 (PRDX2) (ex TDPX1)	208439	SEQ ID No : 183	SEQ ID No : 184	SEQ ID No:314	SEQ ID No:315	SEQ ID No:316
GRB7	124	growth factor receptor-bound protein 7 (GRB7)	130323	SEQ ID No : 185	SEQ ID No : 186	SEQ ID No:79	SEQ ID No:80	SEQ ID No:81
RBBP7	125	retinoblastoma-binding protein 7 (RBBP7)	210874	SEQ ID No : 187	SEQ ID No : 188	SEQ ID No:319	SEQ ID No:320	SEQ ID No:321
TIMP1	126	tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor) (TIMP1)	162246	SEQ ID No : 189	SEQ ID No : 190	SEQ ID No:223	SEQ ID No:224	SEQ ID No:225
YES1	127	v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1 (YES1)	204634	SEQ ID No : 191		SEQ ID No:303	0	SEQ ID No:304
RNF5	128	ring finger protein 5 (RNF5)	112098		SEQ ID No : 192	0	SEQ ID No:25	SEQ ID No:26
PRKCSH	129	protein kinase C substrate 80K-H (PRKCSH)	187232		SEQ ID No : 193	0	SEQ ID No:263	SEQ ID No:264

Symbole gène	N°	Nom	Image	Seq3' US PROV LISTING	Seq5' US PROV LISTING	Seq3' PCT Listing	Seq5' PCT Listing	(mRNA) PCT Listing
CTSD	130	cathepsin D (lysosomal aspartyl protease) (CTSD)	149401	SEQ ID No : 194	SEQ ID No : 195	SEQ ID No:126	SEQ ID No:127	SEQ ID No:128
NEO1	131	neogenin (chicken) homolog 1 (NEO1)	188380		SEQ ID No : 196	0	SEQ ID No:269	SEQ ID No:270
GAPD-a	132	glyceraldehyde-3-phosphate dehydrogenase (GAPD)	152847	SEQ ID No : 197		SEQ ID No:150	SEQ ID No:151	SEQ ID No:152
ACTG1	133	actin, gamma 1 (ACTG1)	182291	SEQ ID No : 198	SEQ ID No : 199	SEQ ID No:242	SEQ ID No:243	SEQ ID No:244
ITGA6	134	integrin, alpha 6 (ITGA6)	182431	SEQ ID No : 200	SEQ ID No : 201	SEQ ID No:245	SEQ ID No:246	SEQ ID No:247
GAPD-b	135	glyceraldehyde-3-phosphate dehydrogenase (GAPD)	153607	SEQ ID No : 202	SEQ ID No : 203	SEQ ID No:166	SEQ ID No:167	SEQ ID No:152
ETV5-b	136	ets variant gene 5 (ets-related molecule) (ETV5)	203394	SEQ ID No : 204	SEQ ID No : 205	SEQ ID No:298	SEQ ID No:299	SEQ ID No:300
CDK4-b	137	cyclin-dependent kinase 4 (CDK4)	195800	SEQ ID No : 206	SEQ ID No : 207	SEQ ID No:286	SEQ ID No:287	SEQ ID No:288
FOS-b	138	v-fos FBJ murine osteosarcoma viral oncogene homolog (FOS)	363796	SEQ ID No : 208	SEQ ID No : 209	SEQ ID No:404	SEQ ID No:405	SEQ ID No:318
HOXA5	139	homeobox protein (HOX-1.3) (ex Hox A5)	300564	SEQ ID No : 210	SEQ ID No : 211	SEQ ID No:382	SEQ ID No:383	SEQ ID No:384
RELA	140	NF-kappa-B transcription factor p65 DNA binding subunit (ex RELa)	122056	SEQ ID No : 212		SEQ ID No:42	0	SEQ ID No:43
SUI1	141	S100 calcium-binding protein A11 (calgizzarin) (S100A11)	155345	SEQ ID No : 213	SEQ ID No : 214	SEQ ID No:186	0	0
ANG	142	angiogenin, ribonuclease, RNase A family, 5 (ANG)	156720		SEQ ID No : 215	0	SEQ ID No:194	SEQ ID No:195
ITGA6	143	integrin, alpha 6 (ITGA6)	182431	SEQ ID No : 216	SEQ ID No : 217	SEQ ID No:245	SEQ ID No:246	SEQ ID No:247
PRMT2	144	HMT1 (hnRNP methyltransferase, S. cerevisiae)-like 1 (HRMT1L1) (ex PRMT2)	158038	SEQ ID No : 218	SEQ ID No : 219	SEQ ID No:201	SEQ ID No:202	SEQ ID No:203
EST R55460	145	EST R55460	154997		SEQ ID No : 220	0	SEQ ID No:185	0
GZMA	146	granzyme A (granzyme 1, cytotoxic T-lymphocyte-associated serine esterase 3) (GZMA)	356763	SEQ ID No : 221	SEQ ID No : 222	SEQ ID No:402	0	SEQ ID No:403
SOX9	147	SRY (sex-determining region Y)-box 9 (campomelic dysplasia, autosomal sex-reversal) (SOX9)	323948	SEQ ID No : 223		SEQ ID No:394	0	SEQ ID No:395
SRF	148	serum response factor (c-fos serum response element-binding transcription factor) (SRF)	321329		SEQ ID No : 224	SEQ ID No:391	SEQ ID No:392	SEQ ID No:393
EDN1	149	endothelin 1 (EDN1)	153424	SEQ ID No : 225		#N/A	#N/A	#N/A
PTPN6	150	protein tyrosine phosphatase, non-receptor type 6 (PTPN6)	66778	SEQ ID No : 226		#N/A	#N/A	#N/A
TFAP4	151	transcription factor AP-4 (activating	159093	SEQ ID No : 227		0	SEQ ID No:210	SEQ ID No:211

Symbole gène	N°	Nom	Image	Seq3' US PROV LISTING	Seq5' US PROV LISTING		Seq3' PCT Listing	Seq5' PCT Listing	(mRNA) PCT Listing
		enhancer binding protein 4) (TFAP4)							
ELF1	152	Human cis-acting sequence.Elf-1	182007	SEQ ID No : 228			SEQ ID No:437	0	0
CD2	153	CD2 antigen (p50), sheep red blood cell receptor (CD2)	120649	SEQ ID No : 229			SEQ ID No:431	0	0
CCND2	154	cyclin D2 (CCND2)	175256	SEQ ID No : 230			#N/A	#N/A	#N/A
IL3RA	155	interleukin 3 receptor (hIL-3Ra)	183087	SEQ ID No : 231			SEQ ID No:440	SEQ ID No:441	0
JUP	156	junction plakoglobin (JUP)	157958	SEQ ID No : 232			#N/A	#N/A	#N/A
RBL2	157	retinoblastoma-like 2 (p130) (RBL2)	108571	SEQ ID No : 233			SEQ ID No:430	0	0
HOXA4	158	homeo box A4 (HOXA4)	110731	SEQ ID No : 234			SEQ ID No:20	SEQ ID No:21	0
ACY1	159	aminoacylase 1 (ACY1)	160764	SEQ ID No : 235			SEQ ID No:435	SEQ ID No:436	0
GADD45A	160	growth arrest and DNA-damage-inducible, alpha (GADD45A)	115176	SEQ ID No : 236			#N/A	#N/A	#N/A
nm23	161	non-metastatic cells 1, protein (NM23A) expressed in (NME1)	174388	SEQ ID No : 237			#N/A	#N/A	#N/A
BBC1	162	ribosomal protein L13 (RPL13) (ex BBC1)	178317	SEQ ID No : 238			#N/A	#N/A	#N/A
VEGFB	163	vascular endothelial growth factor B (VEGFB)	162499	SEQ ID No : 239			#N/A	#N/A	#N/A
LAMR1	164	laminin receptor 1 (67kD, ribosomal protein SA) (LAMR1)	199837	SEQ ID No : 240			#N/A	#N/A	#N/A
IL2RB	165	interleukin 2 receptor, beta (IL2RB)	139073	SEQ ID No : 241	SEQ ID No : 242		SEQ ID No:97	SEQ ID No:98	SEQ ID No:99
DES	166	desmin	153854	SEQ ID No : 243			SEQ ID No:168	SEQ ID No:169	SEQ ID No:170
PRL	167	prolactin	133738	SEQ ID No : 244			SEQ ID No:91	SEQ ID No:92	SEQ ID No:93
CSH1	168	Chorionic somatomammotropin hormone 1 (placental lactogen) = LACTOGEN Precursor	133891		SEQ ID No : 245		SEQ ID No:432	0	0
TEK	169	tyrosine proteine kinase receptor	151501	SEQ ID No : 246	SEQ ID No : 247		SEQ ID No:138	SEQ ID No:139	SEQ ID No:140
Nrg1	170	neuregulin 1 (EST R72075)	155716	SEQ ID No : 248	SEQ ID No : 249		SEQ ID No:189	SEQ ID No:190	SEQ ID No:191
PLAT	rien	pas d'EST ni mRNA	160149				SEQ ID No:433	SEQ ID No:434	0
EST AW184517	rien		image ?						

References

1. DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. (1996) Use of a cDNA microarray to analyze gene expression patterns in human cancer. Nat Genet ,14, 457-460.

2. Jordan, B. R. (1998) Large-scale expression measurement by hybridization methods: from high- density membranes to "DNA chips". J Biochem (Tokyo) ,124, 251-258.

3. Nguyen, C., Rocha, D., Granjeaud, S., Baldit, M., Bernard, K., Naquet, P., and Jordan, B. R. (1995) Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. Genomics ,29, 207-216.

4. Bertucci, F., Van Hulst, S., Bernard, K., Loriod, B., Granjeaud, S., Tagett, R., Starkey, M., Nguyen, C., Jordan, B., and Birnbaum, D. (1999) Expression scanning of an array of growth control genes in human tumor cell lines. Oncogene ,18, 3905-3912.

5. Bertucci, F., Bernard, K., Loriod, B., Chang, Y. C., Granjeaud, S., Birnbaum, D., Nguyen, C., Peck, K., and Jordan, B. R. (1999) Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples [In Process Citation]. Hum Mol Genet ,8, 1715-1722.

6. Ross, J. S. and Fletcher, J. A. (1999) The HER-2/neu oncogene: prognostic factor, predictive factor and target for therapy. Semin Cancer Biol ,9, 125-138.

7. Scorilas, A., Trangas, T., Yotis, J., Pateras, C., and Talieri, M. (1999) Determination of c-myc amplification and overexpression in breast cancer patients: evaluation of its prognostic value against c-erbB-2, cathepsin-D and clinicopathological characteristics using

univariate and multivariate analysis. Br J Cancer ,81, 1385-1391.

8. Fox, S. B., Smith, K., Hollyer, J., Greenall, M., Hastrich, D., and Harris, A. L. (1994) The epidermal growth factor receptor as a prognostic marker: results of 370 patients and review of 3009 patients. Breast Cancer Res Treat ,29, 41-49.

9. Heimann, R., Lan, F., McBride, R., and Hellman, S. (2000) Separating favorable from unfavorable prognostic markers in breast cancer: the role of E-cadherin. Cancer Res, 60, 298-304.

10. Guerin, M., Sheng, Z. M., Andrieu, N., and Riou, G. (1990) Strong association between c-myb and oestrogen-receptor expression in human breast cancer. Oncogene ,5, 131-135.

11. Lim, K. C., Lakshmanan, G., Crawford, S. E., Gu, Y., Grosveld, F., and Douglas Engel, J. (2000) Gata3 loss leads to embryonic lethality due to noradrenaline deficiency of the sympathetic nervous system. Nat Genet ,25, 209-212.

12. Mills, K. J., Vollberg, T. M., Nervi, C., Grippo, J. F., Dawson, M. I., and Jetten, A. M. (1996) Regulation of retinoid-induced differentiation in embryonal carcinoma PCC4.aza1R cells: effects of retinoid-receptor selective ligands. Cell Growth Differ ,7, 327-337.

13. Easty, D. J., Hill, S. P., Hsu, M. Y., Fallowfield, M. E., Florenes, V. A., Herlyn, M., and Bennett, D. C. (1999) Up-regulation of ephrin-A1 during melanoma progression. Int J Cancer ,84, 494-501.

14. Shim, C., Zhang, W., Rhee, C. H., and Lee, J. H. (1998) Profiling of differentially expressed genes in human primary cervical cancer by complementary DNA expression array. Clin Cancer Res ,4, 3045-3050.

15. Tsou, A. P., Wu, K. M., Tsen, T. Y., Chi, C. W., Chiu, J. H., Lui, W. Y., Hu, C. P., Chang, C., Chou, C. K., and Tsai, S. F. (1998) Parallel hybridization analysis of multiple protein kinase genes: identification of gene expression patterns characteristic of human hepatocellular carcinoma. Genomics ,50, 331-340.

16. Schummer, M., Ng, W. V., Bumgarner, R. E., Nelson, P. S., Schummer, B., Bednarski, D. W., Hassell, L., Baldwin, R. L., Karlan, B. Y., and Hood, L. (1999) Comparative hybridization of an array of 21,500 ovarian cDNAs for the discovery of genes overexpressed in ovarian carcinomas. Gene ,238, 375-385.

17. Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D., and Levine, A. J. (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci U S A ,96, 6745-6750.

18. Moch, H., Schraml, P., Bubendorf, L., Mirlacher, M., Kononen, J., Gasser, T., Mihatsch, M. J., Kallioniemi, O. P., and Sauter, G. (1999) High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma. Am J Pathol ,154, 981-986.

19. Rhee, C. H., Hess, K., Jabbur, J., Ruiz, M., Yang, Y., Chen, S., Chenchik, A., Fuller, G. N., and Zhang, W. (1999) cDNA expression array reveals heterogeneous gene expression profiles in three glioblastoma cell lines. Oncogene ,18, 2711-2717.

20. Huang, F., Adelman, J., Jiang, H., Goldstein, N. I., and Fisher, P. B. (1999) Identification and temporal expression pattern of genes modulated during irreversible growth arrest and terminal differentiation in human melanoma cells. Oncogene ,18, 3546-3552.

21. Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., and Sondak, V. (2000) Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature ,406, 536-540.

22. Khan, J., Simon, R., Bittner, M., Chen, Y., Leighton, S. B., Pohida, T., Smith, P. D., Jiang, Y., Gooden, G. C., Trent, J. M., and Meltzer, P. S. (1998) Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. Cancer Res ,58, 5009-5013.

23. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science ,286, 531-537.

24. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., and Staudt, L. M. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling [In Process Citation]. Nature ,403, 503-511.

25. Hoch, R. V., Thompson, D. A., Baker, R. J., and Weigel, R. J. (1999) GATA-3 is expressed in association with estrogen receptor in breast cancer. Int J Cancer ,84, 122-128.

26. Hilsenbeck, S. G., Friedrichs, W. E., Schiff, R., O'Connell, P., Hansen, R. K., Osborne, C. K., and Fuqua, S. A. (1999) Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. J Natl Cancer Inst ,91, 453-459.

27. Martin, K. J., Kritzman, B. M., Price, L. M., Koh, B., Kwan, C. P., Zhang, X., Mackay, A., O'Hare, M. J., Kaelin, C. M., Mutter, G. L., Pardee, A. B., and Sager, R. (2000) Linking gene expression patterns to therapeutic groups in breast cancer. Cancer Res ,60, 2232-2238.

28. Yang, G. P., Ross, D. T., Kuang, W. W., Brown, P. O., and Weigel, R. J. (1999) Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. Nucleic Acids Res ,27, 1517-1523.

29. Perou, C. M., Jeffrey, S. S., van de Rijn, M., Rees, C. A., Eisen, M. B., Ross, D. T., Pergamenschikov, A., Williams, C. F., Zhu, S. X., Lee, J. C., Lashkari, D., Shalon, D., Brown, P. O., and Botstein, D. (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci U S A ,96, 9212-9217.

30. Nacht, M., Ferguson, A. T., Zhang, W., Petroziello, J. M., Cook, B. P., Gao, Y. H., Maguire, S., Riley, D., Coppola, G., Landes, G. M., Madden, S. L., and Sukumar, S. (1999) Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. Cancer Res ,59, 5464-5470.

31. Sgroi, D. C., Teng, S., Robinson, G., LeVangie, R., Hudson, J. R., Jr., and Elkahloun, A. G. (1999) In vivo gene expression profile analysis of human breast cancer progression. Cancer Res ,59, 5656-5661.

32. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R.,

Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O., and Botstein, D. (2000) Molecular portraits of human breast tumours. Nature ,406, 747-752.

33. Hahnel, E., Harvey, J. M., Joyce, R., Robbins, P. D., Sterrett, G. F., and Hahnel, R. (1993) Stromelysin-3 expression in breast cancer biopsies: clinico-pathological correlations. Int J Cancer ,55, 771-774.

34. Skoog, L., Humla, S., Klintenberg, C., Pasqual, M., and Wallgren, A. (1985) Receptors for retinoic acid and retinol in human mammary carcinomas. Eur J Cancer Clin Oncol ,21, 901-906.

35. Thor, A. D., Moore, D. H., II, Edgerton, S. M., Kawasaki, E. S., Reihsaus, E., Lynch, H. T., Marcus, J. N., Schwartz, L., Chen, L. C., Mayall, B. H., and et al. (1992) Accumulation of p53 tumor suppressor gene protein: an independent marker of prognosis in breast cancers. J Natl Cancer Inst ,84, 845-855.

36. Allred, D. C., Harvey, J. M., Berardo, M., and Clark, G. M. (1998) Prognostic and predictive factors in breast cancer by immunohistochemical analysis. Mod Pathol ,11, 155-168.

37. Spencer, K. S., Graus-Porta, D., Leng, J., Hynes, N. E., and Klemke, R. L. (2000) ErbB2 is necessary for induction of carcinoma cell invasion by ErbB family receptor tyrosine kinases. J Cell Biol ,148, 385-397.

38. Behrens, J. (1993) The role of cell adhesion molecules in cancer invasion and metastasis. Breast Cancer Res Treat ,24, 175-184.

39. Roberts, D. D. (1996) Regulation of tumor growth and metastasis by thrombospondin-1. Faseb J ,10, 1183-1191.

40. Taylor-Papadimitriou, J., Burchell, J., Miles, D. W., and Dalziel, M. (1999) MUC1 and cancer. Biochim Biophys Acta ,1455, 301-313.

41. Sneath, R. J. and Mangham, D. C. (1998) The normal structure and function of CD44 and its role in neoplasia. Mol Pathol ,51, 191-200.

42. Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C. F., Trent, J. M., Staudt, L. M., Hudson, J., Jr., Boguski, M. S., Lashkari, D., Shalon, D., Botstein, D., and Brown, P. O. (1999) The transcriptional program in the response of human fibroblasts to serum. Science ,283, 83-87.

43. Theillet, C., Adelaide, J., Louason, G., Bonnet-Dorion, F., Jacquemier, J., Adnane, J., Longy, M., Katsaros, D., Sismondi, P., Gaudray, P., and et al. (1993) FGFRI and PLAT genes and DNA amplification at 8p12 in breast and ovarian cancers. Genes Chromosomes Cancer ,7, 219-226.

44. Granjeaud, S., Nguyen, C., Rocha, D., Luton, R., and Jordan, B. R. (1996) From hybridization image to numerical values: a practical, high throughput quantification system for high density filter hybridizations. Genet Anal ,12, 151-162.

45. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A ,95, 14863-14868.

46. Ferrari, S., Battini, R., and Cossu, G. (1990) Differentiation-dependent expression of apolipoprotein A-I in chicken myogenic cells in culture. Dev Biol ,140, 430-436.

CLAIMS

1. A polynucleotide library useful in the molecular characterization of a carcinoma, said library comprising a pool of polynucleotide sequences or subsequences thereof wherein said sequences or subsequences are either underexpressed or overpressed in tumor cells, further wherein said sequences or subsequences correspond substantially to any of the polynucleotide sequences set forth in any of SEQ ID Nos: 1 - 468 or the complement thereof.

2. A polynucleotide library according to Claim 1 wherein said polynucleotide sequences or subsequences thereof of said pool correspond to any combination of at least one polynucleotide selected among those included in anyone of the following predefined sets :

SET 1: (SEQ ID No:1; SEQ ID No:2); SET 2: (SEQ ID No:3; SEQ ID No:4); SET 3: (SEQ ID No:5; SEQ ID No:6); SET 4: (SEQ ID No:7;SEQ ID No:8); SET 5: (SEQ ID No:9; SEQ ID No:10); SET 6: (SEQ ID No:11: SEQ ID No:12); SET 7: (SEQ ID No:13; SEQ ID No:14;SEQ ID No:15); SET 8: (SEQ ID No:16); SET 9: (SEQ ID No:17; SEQ ID No:18; SEQ ID No:19); SET 10: (SEQ ID No:20; SEQ ID No:21); SET 11: (SEQ ID No:22; SEQ ID No:23; SEQ ID No:24); SET 12: (SEQ ID No:25; SEQ ID No:26); SET 13: (SEQ ID No:27; SEQ ID No:28; SEQ ID No:29); SET 14: (SEQ ID No:30; SEQ ID No:31); SET 15: (SEQ ID No:32; SEQ ID No:33; SEQ ID No:34)) ; SET 16 : (SEQ ID No:35) ; SET 17 : (SEQ ID No:36; SEQ ID No:37; SEQ ID No:38) ; SET 18 : (SEQ ID No:39; SEQ ID No:40; SEQ ID No:41) ; SET 19 : (SEQ ID No:42; SEQ ID No:43) ; SET 20 : (SEQ ID No:44; SEQ ID No:45) ; SET 21 : (SEQ ID No:46; SEQ ID No:47) ; SET 22 : (SEQ ID No:48; SEQ ID No:49; SEQ ID No:50) ; SET 23 : (SEQ ID No:51; SEQ ID No:52; SEQ ID No:53) ; SET 24: (SEQ ID No:54; SEQ ID No:55; SEQ ID No:56) ; SET 25: (SEQ ID No:57; SEQ ID No:58) ; SET 26: (SEQ ID No:59; SEQ ID No:60; SEQ ID No:61) ; SET 27: (SEQ ID No:62; SEQ ID No:63; SEQ ID No:64) ; SET 28: (SEQ ID No:65; SEQ ID No:66; SEQ ID No:67) ;

SET 29: (SEQ ID No:68; SEQ ID No:69; SEQ ID No:70) ; SET 30: (SEQ ID No:71; SEQ ID No:72) ; SET 31 : (SEQ ID No:73; SEQ ID No:74; SEQ ID No:75) ; SET 32 : (SEQ ID No:76; SEQ ID No:77; SEQ ID No:78) ; SET 33 : (SEQ ID No:79; SEQ ID No:80; SEQ ID No:81) ; SET 34: (SEQ ID No:82; SEQ ID No:83) ; SET 35: (SEQ ID No:84; SEQ ID No:85) ; SET 36: (SEQ ID No:86; SEQ ID No:87) ; SET 37: (SEQ ID No:88; SEQ ID No:89; SEQ ID No:90) ; SET 38: (SEQ ID No:91; SEQ ID No:92; SEQ ID No:93) ; SET 39: (SEQ ID No:94; SEQ ID No:95; SEQ ID No:96) ; SET 40: (SEQ ID No:97; SEQ ID No:98; SEQ ID No:99) ; SET 41: (SEQ ID No:100; SEQ ID No:101; SEQ ID No:78) ; SET 42: (SEQ ID No:102; SEQ ID No:103) ; SET 43: (SEQ ID No:104; SEQ ID No:105) ; SET 44: (SEQ ID No:106; SEQ ID No:107; SEQ ID No:108) ; SET 45: (SEQ ID No:109; SEQ ID No:110) ; SET 46: (SEQ ID No:111; SEQ ID No:112; SEQ ID No:113) ; SET 47: (SEQ ID No:114) ; SET 48: (SEQ ID No:115; SEQ ID No:116; SEQ ID No:117) ; SET 49: (SEQ ID No:118; SEQ ID No:119) ; SET 50: (SEQ ID No:120; SEQ ID No:121) ; SET 51: (SEQ ID No:122; SEQ ID No:78) ; SET 52: (SEQ ID No:123; SEQ ID No:124; SEQ ID No:125) ; SET 53: (SEQ ID No:126; SEQ ID No:127; SEQ ID No:128) ; SET 54: (SEQ ID No:129; SEQ ID No:130) ; SET 55: (SEQ ID No:131; SEQ ID No:132) ; SET 56: (SEQ ID No:133; SEQ ID No:134) ; SET 57: (SEQ ID No:135; SEQ ID No:136; SEQ ID No:137) ; SET 58: (SEQ ID No:138; SEQ ID No:139; SEQ ID No:140) ; SET 59: (SEQ ID No:141; SEQ ID No:142; SEQ ID No:143) ; SET 60: (SEQ ID No:144; SEQ ID No:145; SEQ ID No:146) ; SET 61: (SEQ ID No:147; SEQ ID No:148; SEQ ID No:149) ; SET 62: (SEQ ID No:150; SEQ ID No:151; SEQ ID No:152) ; SET 63: (SEQ ID No:153; SEQ ID No:154; SEQ ID No:155) ; SET 64: (SEQ ID No:156; SEQ ID No:157; SEQ ID No:158) ; SET 65: (SEQ ID No:159; SEQ ID No:160; SEQ ID No:161) ; SET 66: (SEQ ID No:162; SEQ ID No:163) ; SET 67: (SEQ ID No:164; SEQ ID No:165) ; SET 68: (SEQ ID No:166; SEQ ID No:167; SEQ ID No:152) ; SET 69: (SEQ ID No:168; SEQ ID No:169; SEQ ID No:170) ; SET 70: (SEQ ID No:171; SEQ ID No:172) ; SET 71: (SEQ ID No:173; SEQ ID No:174; SEQ ID No:175) ; SET 72: (SEQ ID No:176; SEQ ID No:177) ; SET 73: (SEQ ID No:178; SEQ ID No:179) ; SET 74: (SEQ ID No:180; SEQ ID No:181; SEQ ID No:182) ; SET 75: (SEQ ID No:183; SEQ ID No:184) ; SET 76: (SEQ ID No:185) ; SET 77: (SEQ ID No:186) ; SET 78: (SEQ ID No:187; SEQ ID No:188) ; SET 79: (SEQ ID No:189;

SEQ ID No:190; SEQ ID No:191) ; SET 80: (SEQ ID No:192; SEQ ID No:193) ; SET 81: (SEQ ID No:194; SEQ ID No:195) ; SET 82: (SEQ ID No:196; SEQ ID No:197; SEQ ID No:198) ; SET 83: (SEQ ID No:199; SEQ ID No:200) ; SET 84: (SEQ ID No:201; SEQ ID No:202; SEQ ID No:203) ; SET 85: (SEQ ID No:204; SEQ ID No:205) ; SET 86: (SEQ ID No:206; SEQ ID No:207) ; SET 87: (SEQ ID No:208; SEQ ID No:209) ; SET 88: (SEQ ID No:210; SEQ ID No:211) ; SET 89: (SEQ ID No:212; SEQ ID No:213) ; SET 90: (SEQ ID No:214; SEQ ID No:215) ; SET 91: (SEQ ID No:216; SEQ ID No:217) ; SET 92: (SEQ ID No:218; SEQ ID No:219; SEQ ID No:220) ; SET 93: (SEQ ID No:221; SEQ ID No:222) ; SET 94: (SEQ ID No:223; SEQ ID No:224; SEQ ID No:225) ; SET 95: (SEQ ID No:226; SEQ ID No:227) ; SET 96: (SEQ ID No:228; SEQ ID No:229) ; SET 97: (SEQ ID No:230; SEQ ID No:231; SEQ ID No:232) ; SET 98: (SEQ ID No:233; SEQ ID No:234) ; SET 99: (SEQ ID No:235; SEQ ID No:236; SEQ ID No:237) ; SET 100: (SEQ ID No:238; SEQ ID No:239) ; SET 101: (SEQ ID No:240; SEQ ID No:241) ; SET 102: (SEQ ID No:242; SEQ ID No:243; SEQ ID No:244) ; SET 103: (SEQ ID No:245; SEQ ID No:246; SEQ ID No:247) ; SET 104: (SEQ ID No:248; SEQ ID No:249) ; SET 105: (SEQ ID No:250; SEQ ID No:251; SEQ ID No:252) ; SET 106: (SEQ ID No:253; SEQ ID No:254) ; SET 107: (SEQ ID No:255; SEQ ID No:256) ; SET 108: (SEQ ID No:257; SEQ ID No:258) ; SET 109: (SEQ ID No:259; SEQ ID No:260; SEQ ID No:261) ; SET 110: (SEQ ID No:262; SEQ ID No:200) ; SET 111: (SEQ ID No:263; SEQ ID No:264) ; SET 112: (SEQ ID No:265; SEQ ID No:266) ; SET 113: (SEQ ID No:267; SEQ ID No:268) ; SET 114: (SEQ ID No:269; SEQ ID No:270) ; SET 115: (SEQ ID No:271; SEQ ID No:272) ; SET 116: (SEQ ID No:273; SEQ ID No:274) ; SET 117: (SEQ ID No:275; SEQ ID No:276) ; SET 118: (SEQ ID No:277; SEQ ID No:278) ; SET 119: (SEQ ID No:279; SEQ ID No:280; SEQ ID No:281) ; SET 120: (SEQ ID No:282; SEQ ID No:283; SEQ ID No:276) ; SET 121: (SEQ ID No:284; SEQ ID No:285) ; SET 122: (SEQ ID No:286; SEQ ID No:287; SEQ ID No:288) ; SET 123: (SEQ ID No:289; SEQ ID No:290) ; SET 124: (SEQ ID No:291; SEQ ID No:292) ; SET 125: (SEQ ID No:293; SEQ ID No:294; SEQ ID No:295) ; SET 126: (SEQ ID No:296; SEQ ID No:297) ; SET 127: (SEQ ID No:298; SEQ ID No:299; SEQ ID No:300) ; SET 128: (SEQ ID No:301; SEQ ID No:302; SEQ ID No:288) ; SET 129: (SEQ ID No:303; SEQ ID No:304) ; SET 130: (SEQ ID No:305; SEQ ID No:306;

SEQ ID No:307) ; SET 131: (SEQ ID No:308; SEQ ID No:309; SEQ ID No:310) ; SET 132: (SEQ ID No:311; SEQ ID No:312; SEQ ID No:313) ; SET 133: (SEQ ID No:314; SEQ ID No:315; SEQ ID No:316) ; SET 134: (SEQ ID No:317; SEQ ID No:318) ; SET 135: (SEQ ID No:319; SEQ ID No:320; SEQ ID No:321) ; SET 136: (SEQ ID No:322; SEQ ID No:323) ; SET 137: (SEQ ID No:324; SEQ ID No:325) ; SET 138: (SEQ ID No:326; SEQ ID No:327; SEQ ID No:328) ; SET 139: (SEQ ID No:329; SEQ ID No:330) ; SET 140: (SEQ ID No:331; SEQ ID No:332; SEQ ID No:333) ; SET 141: (SEQ ID No:334; SEQ ID No:335; SEQ ID No:336) ; SET 142: (SEQ ID No:337; SEQ ID No:338; SEQ ID No:117) ; SET 143: (SEQ ID No:339; SEQ ID No:340; SEQ ID No:341) ; SET 144: (SEQ ID No:342; SEQ ID No:343; SEQ ID No:344) ; SET 145: (SEQ ID No:345; SEQ ID No:346) ; SET 146: (SEQ ID No:347; SEQ ID No:348; SEQ ID No:349) ; SET 147: (SEQ ID No:350; SEQ ID No:351) ; SET 148: (SEQ ID No:352; SEQ ID No:353) ; SET 149: (SEQ ID No:354; SEQ ID No:355) ; SET 150: (SEQ ID No:356; SEQ ID No:357) ; SET 151: (SEQ ID No:358; SEQ ID No:359; SEQ ID No:360) ; SET 152: (SEQ ID No:361; SEQ ID No:31) ; SET 153: (SEQ ID No:362; SEQ ID No:363; SEQ ID No:364) ; SET 154: (SEQ ID No:365; SEQ ID No:366; SEQ ID No:367) ; SET 155: (SEQ ID No:368; SEQ ID No:369; SEQ ID No:300) ; SET 156: (SEQ ID No:370; SEQ ID No:371) ; SET 157: (SEQ ID No:372; SEQ ID No:373; SEQ ID No:108) ; SET 158: (SEQ ID No:374; SEQ ID No:375; SEQ ID No:376) ; SET 159: (SEQ ID No:377; SEQ ID No:378; SEQ ID No:379) ; SET 160: (SEQ ID No:380; SEQ ID No:381) ; SET 161: (SEQ ID No:382; SEQ ID No:383; SEQ ID No:384) ; SET 162: (SEQ ID No:385; SEQ ID No:386; SEQ ID No:387) ; SET 163: (SEQ ID No:388; SEQ ID No:389; SEQ ID No:390) ; SET 164: (SEQ ID No:391; SEQ ID No:392; SEQ ID No:393) ; SET 165: (SEQ ID No:394; SEQ ID No:395) ; SET 166: (SEQ ID No:396; SEQ ID No:397; SEQ ID No:398) ; SET 167: (SEQ ID No:399; SEQ ID No:400; SEQ ID No:117) ; SET 168: (SEQ ID No:401) ; SET 169: (SEQ ID No:402; SEQ ID No:403) ; SET 170: (SEQ ID No:404; SEQ ID No:405; SEQ ID No:318) ; SET 171: (SEQ ID No:406; SEQ ID No:407; SEQ ID No:408) ; SET 172: (SEQ ID No:409; SEQ ID No:410; SEQ ID No:411) ; SET 173: (SEQ ID No:412; SEQ ID No:413) ; SET 174: (SEQ ID No:414; SEQ ID No:415; SEQ ID No:416) ; SET 175: (SEQ ID No:417; SEQ ID No:418; SEQ ID No:419) ; SET 176: (SEQ ID No:420; SEQ ID No:421; SEQ ID No:422) ; SET 177: (SEQ ID No:423;

SEQ ID No:424; SEQ ID No:425) ; SET 178: (SEQ ID No:426; SEQ ID No:427; SEQ ID No:428) ; SET 179: (SEQ ID No:429; SEQ ID No:408) ; SET 180: (SEQ ID No:430) ; SET 181: (SEQ ID No:431) ; SET 182: (SEQ ID No:432) ; SET 183: (SEQ ID No:433; SEQ ID No:434) ; SET 184: (SEQ ID No:435; SEQ ID No:436) ; SET 185: (SEQ ID No:437) ; SET 186: (SEQ ID No:438; SEQ ID No:439) ; SET 187: (SEQ ID No:440; SEQ ID No:441) ; SET 188: (SEQ ID No:442) ; SET 189: (SEQ ID No:444) ; SET 190: (SEQ ID No:445) ; SET 191 (SEQ ID No:446 ; SEQ ID No:447) ; SET 192: (SEQ ID No:448) ; SET 193: (SEQ ID No:449) ; SET 194: (SEQ ID No:450): SET 195: (SEQ ID No:451) ; SET 196: (SEQ ID No:452) ; SET 197: (SEQ ID No:453) ; SET 198: (SEQ ID No:454) ; SET 199: (SEQ ID No:455) ; SET 200: (SEQ ID No:456) ; SET 201: (SEQ ID No:457) ; SET 202: (SEQ ID No:458) ; SET 203: (SEQ ID No:459) ; SET 204: (SEQ ID No:460) ; SET 205: (SEQ ID No:461) ; SET 206: (SEQ ID No:462) ; SET 207: (SEQ ID No:463) ; SET 208: (SEQ ID No:464) ; SET 209: (SEQ ID No:465) ; SET 210: (SEQ ID No:466) ; SET 211: (SEQ ID No:467) ; SET 212: (SEQ ID No:468)

3. A polynucleotide library according to Claim 2 wherein said polynucleotide sequences or subsequences thereof of said pool correspond to any combination of at least one polynucleotide selected among those included in at least 50%, preferably 75% and more preferably 100% of the predefined sets.

4. A library according to anyone Claim 1 or 2 wherein the pool of polynucleotide sequences or subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising:

SET 1: (SEQ ID No:1 ; SEQ ID No:2) ; SET 4: (SEQ ID No:7 ; SEQ ID No:8) ; SET 18: (SEQ ID No:39 ; SEQ ID No:40 ; SEQ ID No:41) ; SET 21: (SEQ ID No:46 ; SEQ ID No:47) ; SET 24: (SEQ ID No:54 ; SEQ ID No:55 ; SEQ ID No:56) ; SET 32: (SEQ ID No:76 ; SEQ ID No:77 ; SEQ ID No:78) ; SET 38: (SEQ ID No:91 ; SEQ ID

No:92 ; SEQ ID No:93) ; SET 48: (SEQ ID No:115 ; SEQ ID No:116 ; SEQ ID No:117) ; SET 53: (SEQ ID No:126 ; SEQ ID No:127 ; SEQ ID No:128) ; SET 58: (SEQ ID No:138 ; SEQ ID No:139 ; SEQ ID No:140) ; SET 59: (SEQ ID No:141 ; SEQ ID No:142 ; SEQ ID No:143) ; SET 61: (SEQ ID No:147 ; SEQ ID No:148 ; SEQ ID No:149) ; SET 64: (SEQ ID No:156 ; SEQ ID No:157 ; SEQ ID No:158) ; SET 66: (SEQ ID No:162 ; SEQ ID No:163) ; SET 69: (SEQ ID No:168 ; SEQ ID No:169; SEQ ID No:170) ; SET 73: (SEQ ID No:178; SEQ ID No:179) ; SET 85: (SEQ ID No:204; SEQ ID No:205) ; SET 88: (SEQ ID No:210; SEQ ID No:211) ; SET 91: (SEQ ID No:216; SEQ ID No:217) ; SET 97: (SEQ ID No:230; SEQ ID No:231; SEQ ID No:232) ; SET 104: (SEQ ID No:248; SEQ ID No:249) ; SET 105: (SEQ ID No:250 ; SEQ ID No:251 ; SEQ ID No:252) ; SET 112: (SEQ ID No:265 ; SEQ ID No:266) ; SET 113: (SEQ ID No:267 ; SEQ ID No:268) ; SET 115 ; (SEQ ID No:271 ; SEQ ID No:272) ; SET 131: (SEQ ID No:308 ; SEQ ID No:309 ; SEQ ID No:310) ; SET 132: (SEQ ID No:311 ; SEQ ID No:312 ; SEQ ID No:313) ; SET 134: (SEQ ID No:317 ; SEQ ID No:318) ; SET 137: (SEQ ID No:324 ; SEQ ID No:325) ; SET 145: (SEQ ID No:345 ; SEQ ID No:346) ; SET 147: (SEQ ID No:350 ; SEQ ID No:351) ; SET 155: (SEQ ID No:368 ; SEQ ID No:369 ; SEQ ID No:300) ; SET 175: (SEQ ID No:417 ; SEQ ID No:418 ; SEQ ID No:419) ; SET 180: (SEQ ID No:430) ; SET 181: (SEQ ID No:431) ; SET 182: (SEQ ID No:432) ; SET 185: (SEQ ID No:437) ; SET 187: (SEQ ID No:440 ; SEQ ID No:441,

wherein said sequences are useful in differentiating a normal cell from a cancer cell.

5. A polynucleotide library according to Claim 4 wherein said polynucleotide sequences or subsequences thereof of said pool correspond to any combination of at least one polynucleotide selected among those included in at least 50%, preferably 75% and more preferably 100% of the predefined sets.

6. A polynucleotide library according to Claim 4 wherein the pool of polynucleotide sequences or subsequences

correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising:

SET 32: (SEQ ID No:76 ; SEQ ID No:77 ; SEQ ID No:78) ; SET 73: (SEQ ID No:178 ; SEQ ID No:179) ; SET 131: (SEQ ID No:308 ; SEQ ID No:309 ; SEQ ID No:310) ; SET 145: (SEQ ID No:345 ; SEQ ID No:346) and SET 181: (SEQ ID No:431)

and of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising:

SET 38: (SEQ ID No:91 ; SEQ ID No:92 ; SEQ ID No:93) ; SET 58: (SEQ ID No:138 ; SEQ ID No:139 ; SEQ ID No:140); SET 61: (SEQ ID No:147 ; SEQ ID No:148 ; SEQ ID No:149); SET 69: (SEQ ID No:168 ; SEQ ID No:169 ; SEQ ID No:170) and SET 182: (SEQ ID No:432).

7 A polynucleotide library according to Claim 6 wherein said polynucleotide sequences or subsequences thereof of said pool correspond to any combination of at least one polynucleotide selected among those included in at least 50%, preferably 75% and more preferably 100% of the predefined sets.

8. A library according to anyone Claim 1 or 2 wherein the pool of polynucleotide sequences or subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising:

SET 11: (SEQ ID No:22 ; SEQ ID No:23 ; SEQ ID No:24) ; SET 26: (SEQ ID No:59; SEQ ID No:60 ; SEQ ID No:61) ; SET 32: (SEQ ID No:76; SEQ ID No:77 ; SEQ ID No:78) ; SET 34: (SEQ ID No:82 ; SEQ ID No:83) ; SET 40: (SEQ ID No:97 ; SEQ ID No:98 ; SEQ ID No:99) ; SET 57: (SEQ ID No:135 ; SEQ ID No:136 ;SEQ ID No:137) ; SET 64: (SEQ ID No:156 ; SEQ ID No:157; SEQ ID No:158) ; SET 107: (SEQ ID No:255 ; SEQ ID No:256) ; SET 119: (SEQ ID No:279 ;

SEQ ID No:280 ; SEQ ID No:281) ; SET 136: (SEQ ID No:322 ; SEQ ID No:323) ; SET 140: (SEQ ID No:331 ; SEQ ID No:332 ; SEQ ID No:333) ; SET 141: (SEQ ID No:334; SEQ ID No:335 ; SEQ ID No:336) ; SET 145: (SEQ ID No:345; SEQ ID No:346) ; SET 148: (SEQ ID No:352; SEQ ID No:353) ; SET 149: (SEQ ID No:354 ; SEQ ID No:355) ; SET 162: (SEQ ID No:385; SEQ ID No:386; SEQ ID No:387) ; SET 165: (SEQ ID No:394 ; SEQ ID No:395) ; SET 169: (SEQ ID No:402 ; SEQ ID No:403) ; SET 174: (SEQ ID No:414 ; SEQ ID No:415 ; SEQ ID No:416) and SET 188: (SEQ ID No:442),

wherein said sequences are useful in detecting a hormone sensitive tumor cell

9. A polynucleotide library according to Claim 8 wherein said polynucleotide sequences or subsequences thereof of said pool correspond to any combination of at least one polynucleotide selected among those included in at least 50%, preferably 75% and more preferably 100% of the predefined sets.

10. A library according to Claim 8 wherein the pool of polynucleotide sequences or subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising:

SET 32: (SEQ ID No:76 ; SEQ ID No:77 ; SEQ ID No:78) ; SET 136: (SEQ ID No:322 ; SEQ ID No:323) ; SET 145: (SEQ ID No:345 ; SEQ ID No:346); SET 149: (SEQ ID No:354 ; SEQ ID No:355) and SET 169: (SEQ ID No:402 ; SEQ ID No:403)

and of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising:

SET 11: (SEQ ID No:22 ; SEQ ID No:23 ; SEQ ID No:24) ; SET 40: (SEQ ID No:97 ; SEQ ID No:98 ; SEQ ID No:99); SET 57: (SEQ ID No:135 ; SEQ ID No:136 ; SEQ ID No:137); SET 119: (SEQ ID

No:279; SEQ ID No:280 ; SEQ ID No:281) and SET 174: (SEQ ID No:414 ; SEQ ID No:415 ; SEQ ID No:416)

11. A polynucleotide library according to Claim 10 wherein said polynucleotide sequences or subsequences thereof of said pool correspond to any combination of at least one polynucleotide selected among those included in at least 50%, preferably 75% and more preferably 100% of the predefined sets.

12. A library according to anyone Claim 1 or 2 wherein the pool of polynucleotide sequences or subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising:

SET 8: (SEQ ID No:16) ; SET 11: (SEQ ID No:22 ; SEQ ID No:23 ; SEQ ID No:24) ; SET 18: (SEQ ID No:39 ; SEQ ID No:40 ; SEQ ID No:41) ; SET 25: (SEQ ID No:57 ; SEQ ID No:58) ; SET 32: (SEQ ID No:76 ; SEQ ID No:77 ; SEQ ID No:78) ; SET 34: (SEQ ID No:82 ; SEQ ID No:83) ; SET 40: (SEQ ID No:97 ; SEQ ID No:98 ; SEQ ID No:99) ; SET 49: (SEQ ID No:118 ; SEQ ID No:119) ; SET 57: (SEQ ID No:135 ; SEQ ID No:136 ; SEQ ID No:137) ; SET 91: (SEQ ID No:216 ; SEQ ID No:217) ; SET 100: (SEQ ID No:238 ; SEQ ID No:239) ; SET 105: (SEQ ID No:250 ; SEQ ID No:251: SEQ ID No:252) ; SET 136: (SEQ ID No:322 ; SEQ ID No:323) ; SET 138: (SEQ ID No:326 ; SEQ ID No:327 ; SEQ ID No:328) ; SET 139: (SEQ ID No:329 ; SEQ ID No:330) ; SET 141: (SEQ ID No:334 ; SEQ ID No:335 ; SEQ ID No:336) ; SET 158: (SEQ ID No:374 ; SEQ ID No:375 ; SEQ ID No:376) ; SET 169: (SEQ ID No:402 ; SEQ ID No:403) ; SET 180: (SEQ ID No:430) and SET 186: (SEQ ID No:438 ; SEQ ID No:439),

wherein said sequences are useful in differentiating a tumor with lymph nodes from a tumor without lymph nodes.

13. A polynucleotide library according to Claim 12 wherein said polynucleotide sequences or subsequences thereof of said pool correspond to any combination of at least one polynucleotide selected among those included in at least 50%, preferably 75% and more preferably 100% of the predefined sets.

14. A library according to Claim 12 wherein the pool of polynucleotide sequences or subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising

SET 18: (SEQ ID No:39 ; SEQ ID No:40 ; SEQ ID No:41) ; SET 32: (SEQ ID No:76 ; SEQ ID No:77 ; SEQ ID No:78) ; SET 57: (SEQ ID No:135 ; SEQ ID No:136; SEQ ID No:137); SET 91: (SEQ ID No:216 ; SEQ ID No:217) and SET 105: (SEQ ID No:250 ; SEQ ID No:251 ; SEQ ID No:252)

and of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising:

SET 11: (SEQ ID No:22 ; SEQ ID No:23; SEQ ID No:24) ; SET 40: (SEQ ID No:97; SEQ ID No:98 SEQ ID No:99) ; SET 49: (SEQ ID No:118 ; SEQ ID No:119) ; SET 100: (SEQ ID No:238 ; SEQ ID No:239) and SET 141: (SEQ ID No:334; SEQ ID No:335 ; SEQ ID No:336).

15. A polynucleotide library according to Claim 14 wherein said polynucleotide sequences or subsequences thereof of said pool correspond to any combination of at least one polynucleotide selected among those included in at least 50%, preferably 75% and more preferably 100% of the predefined sets.

16. A library according to anyone of Claims 1 or 2 wherein the pool of polynucleotide sequences or

subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising:

SET 11: (SEQ ID No:22 ; SEQ ID No:23 ; SEQ ID No:24) ; SET 22: (SEQ ID No:48 ; SEQ ID No:49 ; SEQ ID No:50) ; SET 23: (SEQ ID No:51 ; SEQ ID No:52 ; SEQ ID No:53) ; SET 26: (SEQ ID No:59 ; SEQ ID No:60 ; SEQ ID No:61) ; SET 28: (SEQ ID No:65 ; SEQ ID No:66 ; SEQ ID No:67) ; SET 31: (SEQ ID No:73 ; SEQ ID No:74 ; SEQ ID No:75) ; SET 32: (SEQ ID No:76 ; SEQ ID No:77 ; SEQ ID No:78) ; SET 34: (SEQ ID No:82 ; SEQ ID No:83) ; SET 49: (SEQ ID No:118 ; SEQ ID No:119) ; SET 57: (SEQ ID No:135 ; SEQ ID No:136 ; SEQ ID No:137) ; SET 64: (SEQ ID No:156 ; SEQ ID No:157 ; SEQ ID No:158) ; SET 73: (SEQ ID No:178; SEQ ID No:179) ; SET 77: (SEQ ID No:186) ; SET 81: (SEQ ID No:194 ; SEQ ID No:195) ; SET 95: (SEQ ID No:226 ; SEQ ID No:227) ; SET 131: (SEQ ID No:308 ; SEQ ID No:309 ; SEQ ID No:310) ; SET 138: (SEQ ID No:326 ; SEQ ID No:327 ; SEQ ID No:328) ; SET 140: (SEQ ID No:331 ; SEQ ID No:332 ; SEQ ID No:333) ; SET 149: (SEQ ID No:354 ; SEQ ID No:355) ; SET 162: (SEQ ID No:385 ; SEQ ID No:386 ; SEQ ID No:387) ; SET 164: (SEQ ID No:391 ; SEQ ID No:392 ; SEQ ID No:393) ; SET 165: (SEQ ID No:394 ; SEQ ID No:395) and SET 183: (SEQ ID No:433 ; SEQ ID No:434),

wherein said sequences are useful in differentiating antracycline-sensitive tumors from antracycline-insensitive tumors.

17. A polynucleotide library according to Claim 16 wherein said polynucleotide sequences or subsequences thereof of said pool correspond to any combination of at least one polynucleotide selected among those included in at least 50%, preferably 75% and more preferably 100% of the predefined sets.

18. A library according to Claim 16 wherein the pool of polynucleotide sequences or subsequences correspond

substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising

SET N° 32: (SEQ ID No:76; SEQ ID No:77; SEQ ID No:78) ; SET N°136: (SEQ ID No:322 ; SEQ ID No:323) ; SET N° 145: (SEQ ID No:345; SEQ ID No:346) ; SET N° 149: SEQ ID No:354; SEQ ID No:355) ; SET N°169: (SEQ ID No:402 ; SEQ ID No:403)

and of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising:

SET No 11: (SEQ ID No:22; SEQ ID No:23 ; SEQ ID No:24); SET No 40: (SEQ ID No:97 ; SEQ ID No:98 ; SEQ ID No:99) ; SET No 57: (SEQ ID No:135 ; SEQ ID No:136 ; SEQ ID No:137) ; SET No 119: (SEQ ID No:279 ; SEQ ID No:280 ; SEQ ID No:281) ; SET No 174: (SEQ ID No:414 ; SEQ ID No:415; SEQ ID No:416).

19. A polynucleotide library according to Claim 18 wherein said polynucleotide sequences or subsequences thereof of said pool correspond to any combination of at least one polynucleotide selected among those included in at least 50%, preferably 75% and more preferably 100% of the predefined sets.

20. A library according to anynone of Claims 1 or 2 wherein the pool of polynucleotide sequences or subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising

SET No 14 (SEQ ID No:30; SEQ ID No:31) ; SET No 23 (SEQ ID No:51; SEQ ID No:52; SEQ ID No:53) ; SET No 25 (SEQ ID No:57; SEQ ID No:58) ; SET No 27 (SEQ ID No:62; SEQ ID No:63; SEQ ID No:64) ; SET No 28 (SEQ ID No:65; SEQ ID No:66; SEQ ID No:67) ; SET No 32 (SEQ ID No:76; SEQ ID No:77; SEQ ID No:78) ; SET No 39 (SEQ ID No:94; SEQ ID No:95; SEQ ID No:96) ; SET No 41 (SEQ ID

No:100; SEQ ID No:101; SEQ ID No:78) ; SET No 44 (SEQ ID No:106; SEQ ID No:107; SEQ ID No:108) ; SET No 48 (SEQ ID No:115; SEQ ID No:116; SEQ ID No:117) ; SET No 51 (SEQ ID No:122; SEQ ID No:78) ; SET No 64 (SEQ ID No:156; SEQ ID No:157; SEQ ID No:158) ; SET No 81 (SEQ ID No:194; SEQ ID No:195) ; SET No 83 (SEQ ID No:199; SEQ ID No:200) ; SET No 91 (SEQ ID No:216; SEQ ID No:217) ; SET No 96 (SEQ ID No:228; SEQ ID No:229) ; SET No 99 (SEQ ID No:235; SEQ ID No:236; SEQ ID No:237) ; SET No 108 (SEQ ID No:257; SEQ ID No:258) ; SET No 110 (SEQ ID No:262; SEQ ID No:200) ; SET No 116 (SEQ ID No:273; SEQ ID No:274) ; SET No 117 (SEQ ID No:275; SEQ ID No:276) ; SET No 118 (SEQ ID No:277; SEQ ID No:278) ; SET No 120 (SEQ ID No:282; SEQ ID No:283; SEQ ID No:276) ; SET No 126 (SEQ ID No:296; SEQ ID No:297;) ; SET No 142 (SEQ ID No:337; SEQ ID No:338; SEQ ID No:117) ; SET No 144 (SEQ ID No:342; SEQ ID No:343; SEQ ID No:344) ; SET No 149 (SEQ ID No:354; SEQ ID No:355) ; SET No 152 (SEQ ID No:361; SEQ ID No:31) ; SET No 153 (SEQ ID No:362; SEQ ID No:363; SEQ ID No:364) ; SET No 154 (SEQ ID No:365; SEQ ID No:366; SEQ ID No:367) ; SET No 157 (SEQ ID No:372; SEQ ID No:373; SEQ ID No:108) ; SET No 159 (SEQ ID No:377; SEQ ID No:378; SEQ ID No:379) ; SET No 162 (SEQ ID No:385; SEQ ID No:386; SEQ ID No:387) ; SET No 166 (SEQ ID No:396; SEQ ID No:397; SEQ ID No:398) ; SET No 167 (SEQ ID No:399; SEQ ID No:400; SEQ ID No:117) ; SET No 168 (SEQ ID No:401) ; SET No 171 (SEQ ID No:406; SEQ ID No:407; SEQ ID No:408) ; SET No 172 (SEQ ID No:409; SEQ ID No:410; SEQ ID No:411) ; SET No 173 (SEQ ID No:412; SEQ ID No:413) ; SET No 176 (SEQ ID No:420; SEQ ID No:421; SEQ ID No:422) ; SET No 177 (SEQ ID No:423; SEQ ID No:424; SEQ ID No:425) ; SET No 178 (SEQ ID No:426; SEQ ID No:427; SEQ ID No:428) ; SET No 179 (SEQ ID No:429; SEQ ID No:408) ; SET No 184 (SEQ ID No:435; SEQ ID No:436) ; SET No 185 (SEQ ID No:437),

wherein said sequences are useful in classifying good and poor prognosis primary breast tumors.

21. A polynucleotide library according to Claim 20 wherein said polynucleotide sequences or subsequences thereof of said pool correspond to any combination of at least one polynucleotide selected among those included in at

least 50%, preferably 75% and more preferably 100% of the predefined sets.

22. A library according to Claim 20 wherein the pool of polynucleotide sequences or subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising

SET N° 23 (SEQ ID No:51 ; SEQ ID No:52 ; SEQ ID No:53) ; SET N° 25 (SEQ ID No:57 ; SEQ ID No:58) ; SET N° 32 (SEQ ID No:76 ; SEQ ID No:77 ; SEQ ID No:78) ; SET N° 41 (SEQ ID No:100 ; SEQ ID No:101 ; SEQ ID No:78) ; SET N° 48 (SEQ ID No:115 ; SEQ ID No:116 ; SEQ ID No:117) ; SET N° 51 (SEQ ID No:122 ; SEQ ID No:78) ; SET N° 64 (SEQ ID No:156 ; SEQ ID No:157 ; SEQ ID No:158) ; SET N° 81 (SEQ ID No:194 ; SEQ ID No:195) ; SET N° 83 (SEQ ID No:199 ; SEQ ID No:200) ; SET N° 91 (SEQ ID No:216 ; SEQ ID No:217) ; SET N° 99 (SEQ ID No:235 ; SEQ ID No:236 ; SEQ ID No:237) ; SET N° 110 (SEQ ID No:262 ; SEQ ID No:200) ; SET N° 116 (SEQ ID No:273 ; SEQ ID No:274) ; SET N° 142 (SEQ ID No:337 ; SEQ ID No:338 ; SEQ ID No:117) ; SET N° 144 (SEQ ID No:342 ; SEQ ID No:343 ; SEQ ID No:344) ; SET N° 149 (SEQ ID No:354 ; SEQ ID No:355) ; SET N° 162 (SEQ ID No:385 ; SEQ ID No:386 ; SEQ ID No:387) ; SET N° 167 (SEQ ID No:399 ; SEQ ID No:400 ; SEQ ID No:117) ; SET N° 171 (SEQ ID No:406 ; SEQ ID No:407 ; SEQ ID No:408) ; SET N° 172 (SEQ ID No:409 ; SEQ ID No:410 ; SEQ ID No:411) ; SET N° 173 (SEQ ID No:412 ; SEQ ID No:413) ; SET N° 176 (SEQ ID No:420 ; SEQ ID No:421 ; SEQ ID No:422) ; SET N° 177 (SEQ ID No:423 ; SEQ ID No:424 ; SEQ ID No:425) ; SET N° 178 (SEQ ID No:426 ; SEQ ID No:427 ; SEQ ID No:428) ; SET N° 179 (SEQ ID No:429 ; SEQ ID No:408) ; SET N° 184 (SEQ ID No:435 ; SEQ ID No:436) ; SET N° 185 (SEQ ID No:437),

and at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising:

SET No 14 (SEQ ID No:30 ; SEQ ID No:31) ; SET No 27 (SEQ ID No:62 ; SEQ ID No:63 ; SEQ ID No:64) ; SET No 28 (SEQ ID